The Effects of Training on Nurses' Performance Using the Abnormal Involuntary Movement Scale

by

Davis C. Clowers, B.S.N.

A Thesis

Presented to

The Oregon Health Sciences University

Department of Psychiatric/Mental Health Nursing

in partial fulfillment of the

requirements for the degree of

Master of Nursing

APPROVED:

[redacted]

Sarah Porter-Tibbetts, R.N., M.P.H., M.S., Assistant Professor

Advisor

[redacted]

Susan J. Will, R.N., M.S., Associate Professor

First Reader

[redacted]

Mary Catherine King, R.N., Psy.D., Associate Professor

Second Reader

[redacted]

Linda Toenniessen, M.D.

Third Reader

[redacted]

Carol A. Lindeman, R.N., Ph.D., Dean, School of Nursing

ACKNOWLEDGEMENTS

There are many people who have helped me finish this project. On the home front, Lilly, my wife, has provided the necessary love, encouragement, prodding, inticements and outright bribery to help me continue. She and my children, Sam and Jenny, have put up with my procrastination. My parents and friends have also given me emotional support and encouragement.

Sarah Porter-Tibbetts, my advisor, has patiently helped me through a series of starts, stops, and changes in direction. Sarah, and the other members of my committee, Mary Kay King, Susan Will and Linda Toenniessen have given of their precious time and energies to me in this project, all the while being very helpful and supportive. Barbara Oyler helped me teach some of the classes.

There are many others, to numerous to name, who have had some part in the completion of this phase of my education. Their contributions are not forgotten.

To these people I wish to say thank you. I hope the fact that I did finish is in some way a demonstration of my appreciation.

To all of the people who insisted that research was fun, I have only one question. Are we having fun yet?

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

Introduction

Since the introduction of the neuroleptic medications in the 1950's, there has been an increase in the number of varieties of neuroleptics, and an increase of the number of patients who take the medications. Unfortunately, the neuroleptic medications used in the treatment of major mental illnesses are also thought to play a large role in the development of an abnormal involuntary movement disorder, the syndrome called tardive dyskinesia (TD). The involuntary movements resulting from TD can create physical, emotional and social problems for the patient. In addition, TD can be irreversible, making it a serious concern to all mental health professionals and mental health patients.

The exact scope of the problem of TD is difficult to define as the reported prevalence of TD tends to vary. Kane et al. (1983) report the prevalence at 14% to 24%, while the American Psychiatric Association Task Force (1980) reported a prevalence rate of 10% to 50%. One reason for the reported variations in prevalence is that the actual presentation of TD fluctuates from time to time. Another reason for varying prevalence rates of TD is that the amount of neuroleptic received by the patient at the time of testing may also affect the presentation of the symptoms. One final reason

variation occurs is because of the method used in detection. For example, a mechanical instrument or a strict count of movements will detect more movements and less severe movement disorders than a subjective observational method.

Predicting prevalence is also difficult. At present the only epidemiological risk factors for TD appear to be the chronic use of neuroleptic medications, advanced age, and gender (elderly females appear to be at higher risk) that help to identify who may develop the disorder (Kane et al. 1983). It is significant, however, that all users of neuroleptic medications are at risk of developing TD.

The problems caused by TD affect not only the users of the neuroleptic medication, but also the professionals who treat the persons and ultimately society at large. For the person, there are physical changes that are readily apparent in public. This can interfere with normal functioning in society. In some severe cases the dysfunction caused by TD can cost people their livelihoods. One of the major goals of the introduction of the neuroleptics was to help individuals return to mainstream society. Ironically, it is possible to successfully treat the psychotic symptoms, yet render the individual unemployable due to the effects on physical appearance and the resulting

handicap of TD.

For the professional, adding to an individual's suffering runs counter to promoting optimal health. Since the neuroleptic medications are not curative and can have serious side effects, there are legal and ethical issues that deserve attention. Gualtiere and Sprague (1984) cite several cases where clinicians have been sued successfully. These cases illustrate the importance of informed consent, careful monitoring and following the American Psychiatric Association (APA) guidelines for the use of neuroleptic medications. These guidelines specify prescribing neuroleptics in the lowest effective dose; using them only when indicated; and examining the patient regularly for signs of TD. The monitoring is easily done and can provide opportunities for the clinician and the patient to negotiate the course of treatment.

In summary, TD is iatrogenic, untreatable, and potentially irreversible. It is not prevented via careful and conscientious monitoring, but such monitoring allows early detection and maximum opportunity for the patient to participate in decisions related to medication risks. Monitoring implies ongoing assessment. However, little attention has been given to how clinicians should be trained to do TD monitoring. One widely used method for monitoring the side effects

of the neuroleptic medications is the Abnormal Involuntary Movement Scale (AIMS). There are several reports in the literature which explore the inter-rater reliability of this monitoring instrument. These reports focus on the inter-rater reliability achieved after training, but the ability of clinicians to use the AIMS without training remains unknown. Since most clinicians do not receive formal training in the use of this instrument, the question of the quality of this monitoring demands exploration.

Statement of Problem

As emphasized earlier, there is an expectation that clinicians, including nurses, screen their patients for TD (Beck, Rawlins & Williams, 1984; Crawford & Kilander, 1985; Critchely & Maurin, 1985; Flaskerud & Van Servellen, 1985; Haber, Leach, Schudy & Sideleau, 1982; Hagerty, 1984; Stuart & Sundeen, 1983; Wilson & Kneisl, 1983). In fact, Nursing has charged itself with the assessment of patients, a process which by definition involves the systematic collection and verification of data (American Nurses' Association, 1982). To accomplish this, the nurse needs valid and reliable instruments with which to perform the assessment, as well as training in how to use there instruments. While training programs tend to increase the reliability of observations and measurement taken, however, they must also be cost effective and efficient. Regarding the use of the AIMS instrument for monitoring TD, it is a fact that, except for a training program available from the National Institute of Mental Health (NIMH) which is time consuming and cumbersome to obtain, there are no widely distributed training programs available to teach clinicians to administer the AIMS. Further, the evidence that does exist regarding the NIMH AIMS

training program's efficiency and effectiveness is ambiguous and does not relate specifically to psychiatric nurses. While the inter-rater reliability studies for the AIMS show high agreement among raters, there are no comparisons between trained and untrained raters. This is an issue because most psychiatric nurses, at least in this author's geographic area, presently using the AIMS are doing so without formal training. This conclusion was reached when the author was unable to locate anyone in the Portland area presently using the AIMS instrument who had had the NIMH training.

This study, therefore, will explore the issue of "training" versus "no training" in the use of the AIMS instrument for monitoring TD by psychiatric nurses. It will go one step further by examining the relationship between experience in psychiatric nursing and ability to use the AIMS instrument. Finally, issues of effectiveness and efficiency of the training model will be explored.

Conceptual Framework

Assessing patients for current or potential health problems for the purpose of affording the patient the best possible outcome is called health screening (Burr & Elwood, 1985). Performing repeated screenings on the same individual is called monitoring. In order to

monitor effectively, nurses must know how to use screening tests properly and efficiently. The conceptual framework for this proposed study depicts how nurses attain proficiency in performing a screening test for TD. The conceptual framework is composed of the learner who strives toward increasing accuracy in TD screening through experience and training.

Training in the use of any screening instrument has been effective when 1) the nurses will be able to perform the screening procedure and evaluate the results accurately and reliably by the end of the training, and 2) the training has been done in an efficient manner. In order to achieve these objectives, the training program must be set up to accommodate the learners at their current level of proficiency and advance them to an acceptable level of proficiency in a short span of time. Since not all nurses begin at the same level of proficiency, it is useful for training program design purposes to examine one way in which nurses are thought to advance from novice to expert in their fields.

Benner (1984) conceptualizes the process of moving from novice to expert as an iterative process of learning and integration of experience. In the beginning, the novice nurse learns rules and procedures which are strictly followed. As these rules and procedures are implemented, however, the nurse is

learning what to observe and how to evaluate the importance of these observations in relation to the individual patient and to other patients who have been similar. Experience with successive patients results in refinement of observational and evaluative skills. As the body of experience grows, the nurse is then able to use an intuitive sense about new patients without the need to follow cumbersome rules. This the "trademark of the expert. What characterizes the "expert" is the ability to see the patient and his or her circumstances holistically rather than as a series or collection of individual occurrences. It follows that the more experiences gained by a nurse (assuming that correct interpretation and refinement occur), the more expert the nurse will become.

Application of Benner's model to this study implies that more experienced, i.e. more "expert," nurses should demonstrate higher levels of ability in screening for TD than less experienced nurses. However, training in the screening process as well as the nurse's reaction to the training process, also influence the accuracy of the results. Figure 1 depicts these factors, all of which are believed to influence the overall progress of the learner.

These three factors (general experience, training and the nurse's reaction to training) are relevant to

this study of nurses' ability to assess for TD. Lane, Glazer, Hansen, Berman and Kramer (1985) introduced the idea that the first of these factors, i.e. general experience of the clinician using the AIMS, gives the rater a sense of what normal client movements are. If this is true, an experienced psychiatric nurse will have a relatively good idea of the movements to expect from medicated patients while a novice psychiatric nurse will not have the same expectations and therefore be less discriminating in assessing movements.

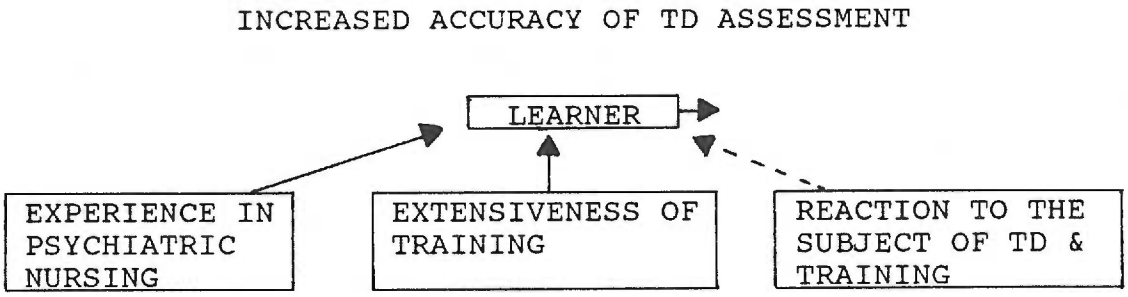
In addition, since the AIMS screening instrument has been in use for more than ten years, it is expected that prior experience with the instrument will be a component of the "general experience" of at least some nurses participating in the study.

The second factor which is conceptualized as driving the learner's progress toward making quality assessments for abnormal movements in clients is the training itself. In practice and as reported in the literature, training in the use of the AIMS consists of programs as minimal as simply reading the directions to as elaborate as a multi-hour training course. In this study, the expectation is that a more detailed and structured training will be more effective than the more abbreviated programs.

The third factor, i.e. emotional reaction to the

training experience, includes the clinician (nurse) reaction to the subject of TD, his or her level of concern regarding the problem of TD, and any other reactions to the training process itself. A positive or negative reaction to the patient, the training experience, or the disease TD itself may enhance or hinder achieving quality assessments.

Figure 1: Conceptual model of the relationship between the accuracy of TD assessments and the learning variables.



For example, there are reports that document the resistance of clinicians to evaluate patients for TD (Brown & Funk, 1986; Munetz, 1985). Several concerns are cited. Fears that litigation will result from evaluating and informing patients about TD are coupled with concerns that patients will elect not to take the prescribed medications. In addition, clinicians are thought to be uncomfortable with discussing iatrogenic illness, which confronts the clinician's beliefs

regarding the value of medication (Munetz,1985). Brown and Funk (1986) are less forgiving toward psychiatry when they cite as one of the reasons for resistance to evaluating patients for TD is that psychiatry tends to minimize and undervalue the patient opinions about the impact of these effects on them.

The clinician may have an avoidance response to evaluating for an iatrogenic illness unless the subject of TD has been examined by the clinician. However, measuring this response to the subject of TD is very complex and would be an entire research project in and of itself. It is mentioned here as one factor thought to have some influence on clinicians' participation in learning about TD, but which will not be addresses in this study as a variable.

Review of the Literature

The following review of the literature will include a definition and description of tardive dyskinesia (TD). Next, to give the reader a fuller appreciation of the screening complexities, categories and testing of TD assessment methods will be discussed. Following that will be a review of the research concerning professional instruction in screening for TD, as the research applies to the conceptual framework. Lastly, issues on videotaped instruction will be addressed, as videotape is the medium of instruction for the only formal training program in TD screening available.

Tardive Dyskinesia

Tardive dyskinesia is a choreiform movement disorder occurring following the long-term use of neuroleptic medications and not explainable as the result of any other neurologic disorder (APA, 1979). At issue scientifically is whether there is a cause and effect relationship between neuroleptic use and TD, but by definition TD follows neuroleptic use. Most basically, no one knows exactly how TD is caused (DeVaugh-Geiss, 1982). But TD is a problem for both patient and practitioner, and all chronically

neuroleptic-exposed patients must be considered at risk.

The clinical presentation of TD is variable. There are two major types of movements which are characteristically observed in the patient with TD. First are the choreic, or jerking movements. Second are the athetoid, or wavelike movements. These movements can be observed occurring alone or in combination. Additionally, sustained contractions (dystonias) may occur, as may tic-like rapid movements. The movements usually occur in the mouth, tongue, face, trunk and limbs of the patient (Jest & Wyatt, 1982; DeVaugh-Geiss,1982).

A patient with TD can present with many movements. The clinician might see mouth movements such as pursing of the lips, thrusting of the tongue, and grinding of the teeth. The patient's head might move from side to side with the neck twisting (torticallis). The patient's arms and hands might make sudden jerky movements or they might move with more writhing motions, with the fingers moving rigidly without apparent purpose. In short, the clinician may see a client moving in a bizarre and purposeless fashion.

In summary, TD is thought to be caused by or closely related to the use of neuroleptic medications. The pattern of movement varies from patient to patient and within a patient over time. This pattern of

manifestation of the syndrome greatly increases the difficulty in performing accurate and definitive TD screening.

Testing of TD Assessment Instruments

Sprague, White, Ullman and Kalachnik (1984) point out that of the several measurement instruments available for TD, none of them have been tested rigorously to assess their measurement qualities. Loney (1980) and Sprague et al. (1984) suggest that measurement instruments should have SCOPE, an acronym for the qualities of systematic, complete, objective, practical and empirical. Because the elements of SCOPE were not defined by the original authors, I will refer to the elements of SCOPE using my own assumptions and interpretations. Systematic will be taken to mean that the test is done in the same way each time. Complete refers to the comprehensiveness of all body parts assessed. Objective is the degree to which observations are made independent of rater bias. Practicality refers to the real life issues of ease of administration and acceptability to the patient. And finally empirical refers to the psychometric qualities of the instrument. The only quality which has been tested to any extent is the empirical quality through inter-rater reliability.

There are three main types of TD assessment

methods: instrumentation, frequency counts, and rating methods. Instrumentation methods use some type of mechanical device to assist in recording movements. An example of this method is the piezoelectric recording device, which measures the bucco-lingual movements with a rubber bulb inserted in the patient's mouth. Frequency counts focus on one or several body parts. Movements per unit of time are counted. Rating methods take into account numerous body movements. Of the rating methods, there are two types, the global rating scales and the multi-item scales. Both scales are intended to rate the patient's movements from every body region. The main differences between the two types of scales are that the multi-item scales are very detailed and require more time to learn and administer than the global scales.

Table 1 compares the three main types of assessment method by the standards of screening theory and SCOPE. Using these criteria, global rating scales are efficient, followed closely by the multi-item scales.

Excluded in this comparison are the disease and treatment characteristics, for these remain the same regardless of the type of assessment method used.

In contrast to the Sprague et al. (1984) criticism that none of the scales had been tested for SCOPE,

TABLE 1
Comparisons of TD Assessment Methods

SCOPE & Screening Theory	Instrumentation (1)	Frequency Counts (2)	Rating Methods Global (3)	Rating Methods Multi-item (4)
Systematic	yes	yes	yes	yes
Complete	no	no	yes	yes
Objective	yes	yes	no	no
Practical	no	yes/no	yes	yes/no
simple	no	yes	yes	no
Rapid App.	no	no	yes	no
Acceptable	no	no	yes	yes
Efficient	no	no	yes	yes
Empirical	yes/no	yes	yes	yes/no
Validity	no	yes	yes	yes
Reliability	yes	yes	yes	no

1 Piezoelectric Method
2 Oral TD Rating Scale, Direct Counting Method
3 AIMS
4 Simpson Abbreviated Dyskinesia Scale

empirical tests were attempted by several researchers. Concurrent validity of the piezoelectric recording was tested by Chien et al. (1980) against five other clinical rating scales. This enabled the researchers to count movements and measure the intensity of movements in the buccolinguo- masticatory area. For concurrent validity, these ratings were compared with the ratings made using five clinical rating scales: the AIMS (Guy, 1976); the Direct Counting Method (Kazamatsuri, Chien & Cole, 1972); the Oral Tardive Dyskinesia Rating Scale (Gerlach, Torsen & Munkuad, 1975); the Crane Rating Scale (Crane & Narango, 1971); and the Sandoz-Wander Tardive Dyskinesia Rating Scale (Simpson, Zoubok & Lee, 1976). Chien et al. (1980) reported concurrent validity correlations ranging from a high of 0.72 to a low of 0.47, for the five scales. Table 2 lists the specific instruments and correlations.

Richardson and Craig (1982) compared the AIMS, a global rating scale, with the Simpson Abbreviated Dyskinesia Scale (ADS), a multi-item scale for both intra- and inter-scale correlations. The total score correlation was significantly high at 0.96, ($p < .0001$). The total facial/oral and total extremities correlations were also high at 0.98 and 0.87 ($p < .0001$) respectively. The intra scale correlations were

TABLE 2

Concurrent Validity Correlations of Five TD Assessment Methods with the Piezoelectric Instrument (Chein et al., 1980)

Scale	Scale Type	Concurrent Validity Correlation
AIMS	Global	0.72 *
Crane Rating Scale	N/A	0.69 *
Oral TD Rating Scale	Frequency Count	0.67 *
Direct Counting Method	Frequency Count	0.62 *
Sandoz-Wander TD Rating Scale	N/A	0.47

* significant at $p < .05$
N/A not available

comparable, with the AIMS correlations tending to be slightly higher than the ADS scores.

While the clinical rating scales for monitoring TD have not been subjected to the rigorous testing suggested by Sprague et al. (1984), some testing of validity and inter-rater reliability has been done. It has been demonstrated that the AIMS compares favorably to the other clinical rating scales in terms of inter-rater reliability and concurrent validity. The AIMS also compares favorably to the other assessment method in relation to SCOPE and the elements of screening theory.

The Chien et al. (1980) study has become the basis for recommending that the AIMS is the clinical rating scale of choice, based upon it's concurrent validity, and it's inter-rater reliability ratings of 0.87 ($p < 0.005$).

Training Programs Using the AIMS

Several research groups have evaluated the AIMS for suitability in clinical use (Germer, Seraydarian & McBrearty, 1984; Lane, Glazer, Hansen, Berman, & Kramer, 1985; Whall et al., 1983). The studies presented will address research in the areas of training program effectiveness, the ability to perform the AIMS accurately, the training program's accommodation to clinicians' levels of experience, and

the learner's reaction to the subject of TD and training for its detection.

Whall et al. (1983) conducted a study to determine the feasibility of a TD screening program by medical/surgical nurses with chronically mentally ill outpatients, anticipating the need for the early detection of TD in a group that will someday become an increased portion of the geriatric population. The researchers trained four, bachelors level, medical-surgical nurses in the use of the AIMS. The training program consisted of preparatory reading, films, video tapes and discussions. The total training time took nine hours for didactic material and discussions. In addition, an unspecified amount of time was used for practice at rating real patients in a neurological disorders clinic. The nurses evaluated 60 outpatient psychiatric patients in teams of two raters. In addition, 12 randomly selected patients were observed by a clinical nurse specialist, described as an expert at assessing abnormal movements, for comparison with the nurses' ratings.

In another study of the development of a training program, Germer et al. (1984) designed a program to instruct psychiatric hospital staff (physicians and registered nurses) in the reliable use of the AIMS. The training program was three hours in length and

utilized the video taped instructions from the National Institute of Mental Health (NIMH). In addition to the regular AIMS procedures, the researchers asked the patients to remove their socks and shoes to facilitate observation of foot movement. They also eliminated the procedure of assigning a lesser severity to movements which were produced by activation. Patients were also asked whether they noticed movements in themselves or others. They made these modifications to enhance the observation of foot movements and to increase the sensitivity in scoring. Group discussions facilitated the corrections in any inconsistent scoring obtained in practice sessions. Also, the trainees read review articles on TD, which took an unspecified amount of time. The inter-rater reliability scoring was done using a video tape of two patients at Haverford hospital.

Rather than study a training program, Lane et al. (1985) conducted a study to compare experienced versus inexperienced raters on their abilities to evaluate patients with TD. The four raters, two psychiatrists experienced with TD and the AIMS, and two psychiatric residents with little clinical exposure to TD, evaluated 33 psychiatric patients and made simultaneous ratings of the patients' movements. The raters alternated conducting the AIMS each week to control for

differences in style of conducting the exams. The data were collected over a ten month period.

All three studies cited demonstrate that the AIMS procedures can be effectively learned by the doctors and nurses who participated in the studies. The authors measured the accuracy of learning through inter-rater reliability scores. Unfortunately, the comparability of the studies is hampered by the variations in training times, clinician discipline and preparation, study design and in the statistical analyses used. However, as can be seen in table 3, the inter-rater reliability in all three studies is high. While all three studies addressed the apparent effectiveness of training, none of the studies addressed the usefulness of the AIMS when the untrained clinician merely reads the instructions before use.

Lane et al. (1985) set out by design to compare experienced versus inexperienced raters using the AIMS. Again, it is unfortunate that the training method was not explicitly described making it difficult to compare the study with Whall et al. (1983) and Germer et al. (1984). Neither of the latter two studies specifically examined the relationship between experience in psychiatry and accurate learning or use of the AIMS. This leaves the area of the effect of experience wide open for speculation. There is, however, inferential

Table 3
TD Screening Training Studies

Study	Type of Instrument	Training Time (hrs)	Clinician Training	Number of Raters	Number of Patients	IRR
Whall et al.(1983)	Global (AIMS)	12+	BSN (non-psych)	4	60	.90 (total score)
Germer et al.(1984)	Global (AIMS + mod.)	3 1/2	RNs Psychiatrist Adm. MDs	15 15 9	2	.97 (IRR coeff.)
Lane et al. (1985)	Global (AIMS)	Unknown	Faculty Psychiatrists Psy. Res.	2 2	33	.79 (ICC) .81 Pearson R

evidence that the Whall study did compensate for non-psychiatric background of the nurses studied by using the most extensive training program. Similarly, the Germer study used a much shortened training program for the staff of a psychiatric hospital.

While Whall et al. (1983) commented on the acceptability of the AIMS in relation to the patient population, none of these researchers reported how the clinicians reacted to the subject of TD or the training program itself. There are other reports in the literature, however, documenting resistance to screening for TD (Brown & Funk, 1986; Munetz, 1985). These suggest that clinicians may be slow to recognize the syndrome and may actually resist adding TD screening to their monitoring repertoire (Brown & Funk, 1986).

In summary, the review of research dealing with training and use of the AIMS examination illustrates that high levels of inter-rater reliability can be achieved. In addition to the training, experience with the TD syndrome and the population at risk may have some influence upon achieving high levels of inter-rater reliability. The comparability of the research is low due to differences of study design, population differences and modifications of the AIMS. Due to the low comparability of the studies, questions arise. How

much training is necessary to achieve high levels of inter-rater reliability? Does experience with TD have much influence upon inter-rater reliability?

Two of the three above mentioned research projects used NIMH training videotapes as well as didactic instruction. Since the training methods established by NIMH involve the use of videotaped instruction, a discussion of the merits and drawbacks of videotaped instruction will follow.

Videotaped Instruction

It is conceivable that the assessment of the symptoms of TD could be learned through the literature and individual instruction. Perhaps the efficiency of such an instruction program could be enhanced by live demonstration before a large group. This depends upon the availability and willingness of a suitable client as well as the availability of the students. Since the 1940's and 50's, the idea of using videotaped instruction has existed as a method for instruction (Gibbons, Kincheloe & Down, 1977). As a technique for instruction, Chu and Schramm (1967) concluded that students can learn as well from television as they can from classroom techniques. The educational establishments counter--charged that the video programs lack the flexibility of the textbook and classroom techniques. In answer to the charges, a technique

known as Tutored Videotape Instruction (TVI) was developed to give students the opportunity to view lectures with standardized content, along with the flexibility to stop the lecture for questions and discussion (Gibbons et al., 1977).

Sox, Marton, Higgins, and Hickam (1984) used the TVI method in studying teaching effectiveness. They compared the pre- and post-test scores of medical students taking a course in clinical decision-making employing the TVI method, and identical live lectures. They found that the students did equally well learning the material using either method. The authors pointed out that when expertise is scarce, the TVI method allows for broad distribution of lectures without the necessity of an expert. The students reported that they felt much more comfortable asking questions and interrupting the tape than they had felt in other lecture situations.

The techniques used by Germer et al. (1985) and Whall et al. (1983) fit within the TVI concept. Readings, videotaped instruction, practice and discussions to answer questions and clarify key points were used.

Summary

TD is a major problem confronting mental health professionals. The syndrome is thought to be

iatrogenic and therefore demands the attention of mental health professionals. An acceptable assessment instrument, the AIMS, has been tested and used widely in the monitoring of TD. It has also been useful in the detection of TD symptoms. It is known that the AIMS can be taught to nurses and other staff so that inter-rater reliability is high. Use of the NIMH training methodology to train nurses to administer the AIMS is supported by a body of research. The projects in which this was done used extensive training programs and included a variety of staff educated at different levels. However, the studies were sufficiently different from one another that one-to-one or across-the-board comparisons are not possible. Additional research is indicated.

Purpose of the Research

The purpose of this research is to examine the questions:

1) To what extent does a structured training program affect the ability of nurses to achieve accurate results using the AIMS?

2) To what extent are these results different from the results of nurses who have read the instructions for AIMS only.

3) To what extent does level of experience affect nurses performance using the AIMS.

Hypotheses

Hypothesis number one (H1): Psychiatric nurses who receive the formal National Institute of Mental Health (NIMH) training will have post-test scores which significantly approximate the normative standard scores from NIMH ($p > 0.05$).

Hypothesis number two (H2): Psychiatric nurses who receive the formal NIMH training will have significantly lower variability on the AIMS (post-test) than psychiatric nurses who have not had formal training.

Hypothesis number three (H3): There will be significant improvement of the experimental group's accuracy from pretest to post-test on the AIMS when formal NIMH training occurs between pretest and post-test.

Hypothesis number four (H4): The post-test scores of the experimental group will be significantly more accurate than the post-test scores of the control group.

Hypothesis number five (H5): Psychiatric nurses who have greater than one year of experience in psychiatric nursing will have pretest scores significantly closer to normative standard scores than psychiatric nurses with less than one year of experience in psychiatric nursing.

Definitions

Accuracy: The degree to which an AIMS score approximates the normative standard score.

Normative Standard Score: An AIMS score (item #8, Severity of abnormal movements) for a specific patient, established by NIMH and included in the NIMH training program.

Psychiatric Nurse: A registered nurse who works directly with psychiatric patients in an inpatient or outpatient setting in which neuroleptics are prescribed.

Training: A program of instruction where the trainee is shown how to perform the AIMS, allowed to practice, and corrections in technique and assessment judgement are made.

Experience: Past and/or current practice at performing a set of skills.

CHAPTER TWO

The following study was designed to test the hypotheses stated in the preceding chapter. The description of the study will include sections on design, sample, setting, protection of human subjects, instruments, data collection procedures, and analysis.

Design

An experimental design of pretest, post-test with a control group was planned. Psychiatric nurses were invited to participate in the study and to have the opportunity to learn to do TD screening using the AIMS. Four training sessions were conducted. In the first and fourth sessions, respondents were randomly assigned to the control or experimental group which ran simultaneously. In the second and third sessions, the conditions of being allowed to do the training precluded randomization. Therefore the respondents in the second session were assigned to the experiemental group and the respondents in the third session were assigned to the control group. When this change was made, randomization of participant assignment stopped, altering the design of the study. It was hoped that no significant differences between the two groups would emerge due to this change. The significance of this change will be discussed in Chapter IV.

The experimental group received the NIMH training

on the AIMS after the pretest and before the post-test (see Appendix F). The control group received the pretest, a lecture about TD without training in the use of the AIMS and a post-test. The design is diagramed as follows:

Experimental	R	O_1	X	O_2
Control	R	O_3		O_4

R= Randomization
O= Measurement
X= Intervention

with O_1 and O_2 representing the experimental pre- and post-tests respectively, O_3 and O_4 representing the pre- and post-tests respectively and X representing the NIMH training.

The changes that occurred in the random assignment the participants into groups changed the design to a quasi-experimental design known as the non-equivalent control group design. The controls for internal and external threats to study validity are similar to the experimental design above, but the weakness that differentiates it from the experimental design is that there is always a question as to whether the control group and the experimental group are alike. The method for comparing the groups is through the comparison of demographic data and comparison prior to the treatment of the pretest scores (Polit & Hungler, 1987).

Methods

Subjects for the first session were notified through fliers sent to psychiatric hospitals and clinics in cities along the Willamette valley from Eugene, Oregon, to Vancouver, Washington. The fliers (Appendix E) stated that free instruction in using the AIMS was to be given and that participation would be used as part of a research study. When the first session produced sparse attendance, the training was offered to hospital staffs where there was a likelihood of having good attendance. The second and third sessions were done at hospital settings. Finally, a fourth session was held for nurses who had expressed an interest in attending the AIMS class but were not able to attend on the first advertised date.

The change in data collection occurred primarily due to the low attendance at the training sessions. The groups were small, less than 7, except for one session where 10 participants attended. With each repeat of the training, it became more difficult to attract participants so data collection was stopped when the number of completed data sets was sufficient to provide statistical validity in testing most of the hypotheses.

As the participants pre-registered for the first session, they were randomly assigned to either the

experimental group or the control group. To assure an even distribution of experienced and non-experienced nurses to the experimental and control groups, the nurses were listed on consecutively numbered sheet. Each consecutive blank had been previously randomly assigned to either the experimental or control group by coin toss. The assignment for the blank determined the assignment of the participant. The groups were separated into two rooms. The first training involved two trainers. One was the researcher; the other was a nurse with several years experience in psychiatric nursing, who had been trained by the researcher to be familiar with the AIMS procedure, the AIMS training and relevant literature on the subject of TD screening. In subsequent training sessions the researcher presented both training formats, although at different times. Consistency of instruction was most important during the period of data collection for the control group. Each group received an explanation of the study and assurance that both groups would receive the same information albeit in a different order of presentation.

Participants also filled out consent forms and demographic questionnaires. For the pretest, both groups were given a copy of the AIMS and instructed to read the instructions on the back of the form. The

subjects then rated pretest/ post-test patient number five (Table 4) from the NIMH training video (the pretest) and the ratings were collected. The NIMH training video was presented to the experimental group, with discussions and practice. At the end of the session, the experimental group rated the same client presented in the pretest. While the experimental group viewed the NIMH training video, a live lecture on TD was presented to the control group (approximately 30 minutes). Following the lecture, the control group took the post-test. Since this was the end of the data collection period for the control group, the NIMH training video was then presented. To make sure that all respondents had access to the same information, the lecture on TD was offered to the experimental group after the post-test.

NIMH AIMS Training Tape

The NIMH training tape is a 40 minute video presentation of how to use the AIMS. It includes a brief explanation of the problem of TD, a description of the AIMS instrument, two demonstrations of how to perform the AIMS examination procedure and four practice sessions with individuals presented as they are being examined via the AIMS examination procedure (The original NIMH tape has five practice examples, but for this study, the fifth individual was isolated on tape

Table 4
Practice Example AIMS Ratings
from the NIMH Training Tape

Example Patient	Highest Rating (1)	Body Areas Affected (2)	NIMH Rating on Item # 8 (3)
#1	4	7	4
#2	4	7	4
#3	3	7	2
#4	2	4	2
#5 Pretest/ Post-test	2	3	2

(1) Highest rating given by NIMH on any single body area (items 1-7), in terms of severity. 0 = none; 1 = minimal; 2 = mild; 3 = moderate; 4 = severe

(2) Number of AIMS items (Items 1-7) affected.

(3) Severity of abnormal movements, rated on the same scale as items 1-7.

for use as the pre and post-test). Each of the practice examples contain the scoring supplied by NIMH.

The demonstrations of the AIMS examination procedure are presented first with an individual without a movement disorder and then with an individual with a very severe movement disorder. Neither of these demonstrations were given ratings. The individuals used in the practice examples were filmed in Japan and probably for language reasons, the tape for these examples is silent. The remaining examples are about 6 to 7 minutes in length, followed by the ratings from NIMH. The first two individuals shown in the examples are severely afflicted with a movement disorder evident in every body area defined by the AIMS. The third and fourth examples exhibit less severe abnormal movements, but the movements are present in several body parts. Table 4 lists the practice example ratings and the pretest/post-test ratings. Time for scoring and discussion of the NIMH ratings was allowed after each practice example.

Sample

The subjects of this study were psychiatric nurses from the Salem and Portland metropolitan areas. Participation was limited to English speaking registered nurses. A minimum number of participants was set at 35, however, data collection was concluded

after 34 sets of data were collected as this number was slightly over the minimum amount needed to perform the statistical analyses proposed for hypostheses 1 through 4. As indicated above, to achieve a significant number of subjects to adequately test hypothesis number 5, trainings would have had to continue for an indefinite period with low probability of attracting many more novice nurses. The major employers of novice psychiatric nurses had already been contacted with few respondents resulting.

Protection of Human Subjects

The AIMS ratings of the subjects were anonymous except for a code number which identified the demographic questionnaire, pretest and post-test as belonging to the same individual as well as to which group the data belonged. A consent form (Appendix C) was also filled out by each participant.

As mentioned above, the subjects for this study were nurses who came to hear a presentation about the use of the AIMS. It was thought to be important that all respondents had the opportunity to receive the information whether they were in the experimental group or the control group. Therefore, all of the NIMH training and the lecture was presented to both groups. The order of presentation was different so that data collection would take place in compliance with the

design of the the study. Appendix F is a schedule of events for each group demonstrating the content and data collection periods for each group.

The training was advertised through letters and announcements (see Appendices D and E) sent to psychiatric hospitals and clinics. Both the letters and announcements clearly stated that the training was part of a research project so the participants would be well aware of the study.

Instruments

The instruments used in this study were the AIMS (Appendix A) and a demographic questionnaire (Appendix B). The AIMS was chosen because it is the only widely available instrument for screening TD that has a widely available training program. It is also the instrument which has been tested with favorable results in several research studies. It is easily administered and has been shown to have high inter-rater reliability. The demographic questionnaire provides data for comparison of the experimental group with the control group and to determine length of experience for hypothesis number five.

Data Analysis

All data analyses were made using the responses from the completed demographic questionnaires, pretests and post-tests. The data were analyzed with the use of

the Crunch statistical program. The following will be descriptions of how each hypothesis was tested.

To begin, the experimental group was compared to the control group to test for the equivalence of the two groups. As this study calls for the sampling of a population, the groups must be comparable for the statistics to be valid (Polit & Hungler, 1987). The means of the participants' ages and number of years in psychiatric nursing were compared. In addition, responses to use of the AIMS in practice and formal training in the use of the AIMS were compared. These comparisons were made using the t-test.

To test hypothesis number one, the post-test scores of the experimental group were compared with the normative standard score from NIMH by way of a t-test. Specifically, item number eight, severity of abnormal movements, was used for comparison. The NIMH score for item number eight was compared with the post-test scores of the experimental group. The level of significance was set at $p > .05$.

Since the AIMS is not a scale in the true sense, that is, totalled or condensed into one score, the only global rating of the instrument is item number eight. The psychometric properties of the AIMS as a scale have not been published. Therefore it is not known to what extent item 8 is an index score for items 1-7. Lane et

al. (1985) suggest that the overall severity score (item number eight) is the single most useful rating of the AIMS for the monitoring of patients over time. Since the nursing role in the early detection of TD is to do the regular monitoring, item number eight was selected as the item on which to compare the participants to the NIMH scores.

Hypothesis two was tested by comparing the post-test scores on item eight for the experimental and control groups. The statistical test used was the F test to measure the variability among participant's scores.

Hypothesis number three was tested using the t-test to compare the experimental group's performance on item eight of the AIMS. It is necessary that there be a directional difference from relative inaccuracy to accuracy for H3 to be accepted. The experimental group's mean pretest score was compared with the post-test score. Significant directional difference from pretest to post-test ($p > .05$) determines the acceptability of H3.

Hypothesis number four was tested by comparing the post-test scores of the experimental and control groups. As the accuracy of the experimental group's post-test scores have been tested in H1, the comparison of the experimental and control groups post-test scores

using the t-test provides the test for significance between the two groups.

Hypothesis number five was to be tested by using a one way analysis of variance (ANOVA). However, there were not enough subjects with less than one year of experience required to test the hypothesis.

Chapter III

Results

The following chapter is the presentation of the results from the data analysis. A description of the experimental and control groups' characteristics will be presented, followed by the data analysis for the individual hypotheses. The results are presented for an N of 34 (experimental group n = 17, control group n = 17) except where specifically noted.

Group Characteristics

As stated in the preceding chapter, the experimental and control groups must be compared on the basis of the participants' mean ages and number of years in psychiatric nursing. Also, the two groups must be compared on the basis of past experience with the AIMS and formal training in the use of the AIMS.

The total sample included nurses who ranged in age from 27 to 73 years, with the mean age being 43 years, 1 month; median of 42 years; and a mode of 39 years. The experimental group had a range of 32 to 60 years, with a mean age of 44 years, 1 month, while the control group had a range of 27 to 73 years, with a mean age of 42 years, 1 month. The two groups are not statistically different when compared by age as computed using the t test ($t = 0.652$ at $df = 32$).

The range of years spent in psychiatric nursing

for the total sample had a range of nine months to 24 years with the mean time being 9 years; median of 9 years; and an even distribution across the range. The experimental group had a range of 1 to 24 years and a mean of 10 years. The control group had a range of 9 months to 19 years, with a mean of 8 years. Again, there was no significant difference between the two groups as determined by a t-test ($t = 1.207$ at $df = 32$). Table 5 summarizes the comparison of key demographic data.

In addition to examining the above group characteristics for differences, the groups' experience with the AIMS and prior training in the use of the AIMS were examined. As shown in table 5, it was determined that the control group had significantly more participants who had used the AIMS than did the experimental group ($p < 0.162$). This difference is important as it relates to the experience factor of the conceptual model and the assumptions of the statistical analysis of the study, i.e. that the groups be equivalent.

The differences between the groups was non-significant in regards to any prior formal training received by the participants. Analysis of the remaining demographic data revealed no statistical differences between the groups. These data items

Table 5

Age and Experience Characteristics of the Experimental and Control Groups

Characteristics	Experimental	Control	Test	Significance
Age (years)*				
Range	32-60	27-73		
Mean	44	42	t-test	ns
SD	7.466	10.380		
Years in psych. nursing *				
Range	1-24	0.75-19		
Mean	10	8	t-test	ns
SD	6.454	5.143		
Use of AIMS clinically				
Yes (freq.)	5	13		
No (freq.)	12	4	x^2	p<0.0162
Formal training with AIMS				
Yes (freq.)	0	4		
No (freq.)	17	13	x^2	ns

* rounded to the nearest year.

include gender, time spent as an RN, the participant's highest level of nursing education, and the number of years spent in specific areas of psychiatric nursing.

Another test of the differences between the two groups is to compare them at the point where one would expect them to be the same, that is, at the pretest. On item number 8, severity of abnormal movements, which is the same item with which the groups were compared to test the hypotheses, the mean scores were not significantly different ($t = 1.21$ at $df = 31.49$, $p > 0.05$). Therefore, by this test the groups appear to be similar. However, this similarity must be observed with caution in that a charateristic not considered in this statistic is the differences between the groups in relation to their prior experience using the AIMS. There is no certainty that the groups will act the same over time even though they are the same at one point in time.

The AIMS item used for comparison for all five hypotheses was number 8, Severity of abnormal movements. An NIMH rating of 2 (mild severity) was given for the videotaped patient used for the pre and post-test.

Findings for Hypothesis Number One

Hypothesis number one states that psychiatric nurses who receive the formal National Institute of

Mental Health (NIMH) training will have post-test scores which significantly approximate the normative standard scores from NIMH ($p > 0.05$). The experimental group rated the videotaped patient. The mean score was 0.824 ($n = 17$, $SD = 0.393$). Since the computed value of t is less than the table value at ã = 0.05, the hypothesis is rejected and the experimental group's score is considered to be significantly different from the NIMH score.

Findings for Hypothesis Number Two

Hypothesis number two states that psychiatric nurses who receive the formal NIMH training will have significantly lower variability on the AIMS (post-test) than psychiatric nurses who have not had formal training. The ratings for the AIMS are 0 = none; 1 = minimal; 2 = mild; 3 = moderate; and 4 = severe. The control group post-test scores varied over four ratings while the experimental group varied over two of the ratings (see table 6).

Using the F test for equal variances, the hypothesis is supported (ã = 0.05) in the respect that the experimental group had lower variability in the post-test scoring. However, it was assumed that the lower variability would occur along with increased accuracy as compared to the NIMH score.

Table 6

AIMS Post-test Score Frequencies on Item # 8
by Control and Experimental Groups

Ratings	Frequencies Control Group	Experimental Group	NIMH
0 None	2	3	
1 Minimal	6	14	
2 Mild	5	0	*
3 Moderate	4	0	
4 Severe	0	0	

Findings for Hypothesis Number Three

Hypothesis number three states that there will be significant improvement of the experimental group's accuracy from pretest to post-test on the AIMS when formal NIMH training occurs between pretest and post-test. The experimental group's mean score dropped from 1.765 on the pretest to 0.824 on the post-test, which is significant at ã = 0.05. The experimental group mean score moved even further away from the NIMH score of 2 on the post-test. Since the experimental group's performance was away from the score of NIMH, the hypothesis is not supported.

Findings for Hypothesis Number Four

Hypothesis number four states that the post-test scores of the experimental group will be significantly more accurate than the post-test scores of the control group. The post-test score mean of the experimental group was 0.824 and the post-test score mean of the control group was 1.647 which when compared to the NIMH score of 2.000 indicates that the experimental group post-test scores were not significantly more accurate than the post-test scores of the control group as measured by a t-test. Therefore, the hypothesis is not supported.

To further understand the experimental and control group's post-test performance, both groups' pretest-to-

post-test performance were also analyzed. The control group started with a mean score of 1.412 on the pretest and moved to a mean score of 1.647 on the post-test. While this is slightly in the direction of matching with the NIMH score of 2, it is not a significant difference (ã = 0.05). Thus, the experimental group's post-test mean of 0.824, moves away from the NIMH score of 2, and the control group's score stays essentially the same, but closer to the NIMH score.

Findings for Hypothesis Number Five

Hypothesis number five states that psychiatric nurses who have greater than one year of experience in psychiatric nursing will have pre-test scores significantly closer to normative standard scores than psychiatric nurses with less than one year of experience in psychiatric nursing. Unfortunately, of the 34 subjects participating in the study, only three subjects had psychiatric nursing experience of one year or less, making statistical analysis for the hypothesis very weak.

The next level of experience that allows for statistical analysis is at psychiatric nursing experience of less than five years, or greater than or equal to five years. An ANOVA of the data revealed interesting results which were nevertheless not significant. The results are presented in Table 7.

Table 7

Analysis of Variance for RN's Performance on
NIMH Pretest Item Number Eight

Group	N	Mean Score Item # 8	SD	n with prior AIMS use
Experimental, Psych experience < 5 years	5	1.400	0.5477	1
Experimental, Psych experience 5 years	12	1.917	0.9962	4
Control, Psych experience < 5 years	6	1.000	0.6235	5
Control, Psych experience 5 years	11	1.636	0.8090	8

Source	df	SS (H)	MSS	F	P
Between Subjects	33	24.2353			
RN Groups	3	3.5732	1.1911	1.729	0.1802
Subj with Groups	30	20.6621	0.6887		

The results presented in Table 7 show a tendency for the more experienced groups of nurses to be closer in accuracy to the NIMH score of 2, than the less experienced nurses. Also included in the table, are the numbers of participants in each group who had used the AIMS clinically. It appears that the prior use of the AIMS was spread somewhat evenly amongst each group. Approximately 55% of the less experienced nurses had prior clinical use of the AIMS and 52% of the more experienced nurses had prior clinical use of the AIMS. If the more experienced nurses had all of the prior experience with the AIMS, then the trend toward more experience producing a more accurate rating would point to AIMS use as the experience factor. As it is, it appears that the trend toward accuracy increases with general experience as a psychiatric nurse. Although perhaps not statistically significant, the trend clearly supports the value of the AIMS training, with both experienced and less experienced nurses.

Chapter IV

Discussion and Conclusions

The purpose of this study was to examine the questions:

1) To what extent does a structured training program affect the ability of nurses to achieve accurate results using the AIMS?

2) To what extent are these results different from the results of nurses who have read the instructions for AIMS only.

3) To what extent does level of experience affect nurses performance using the AIMS.

This chapter discusses the results of the data analysis on these questions. Following this will be a discussion of further questions and hypotheses generated by the findings of the study as well as a discussion of the limitations of the study. The findings and conclusions of the study will be summarized, followed by the implications for nursing practice and recommendations for further study.

Limitations of Study Design

According to Campbell and Stanley (1966), the pretest, post-test randomized control group design controls for all eight internal sources of invalidity. However, the changes that occurred in data collection which ended the randomization of participants into

groups weakened the control for the internal variable of selection. The selection variable is important in the respect that there was a biased selection of participants assigned to the groups (i.e. most participants in the control group had experience using the AIMS while the participants in the experimental group did not).

The planned design provides fair controls for threats to internal validity, however, there are some concerns regarding the external validity or generalizability of the proposed study. The design does not control for the interaction of testing and the variable; and in general, the design provides questionable control for two other sources of external invalidity which are the interaction of selection and the variable; and the reactive arrangements. The aberration which occurred in the randomization of the subjects further weakens the control for interaction of selection and the variable.

The possibility of an interaction between testing and the variable exists in this study. The problem exists when the participants are cued to the material being presented, possibly causing them to pay more attention to the material than they would without a pretest. However, since the pretest is a onetime practice which was presented in the training, it seems

likely that an effect would be an enhancement toward learning. Since pretesting is a common practice in continuing education courses for nurses, it appears reasonable to expect that pretesting would be used in non-research training of nurses and therefore this study is similar to how general training might be conducted.

A similar process, that of demand characteristics, a form of the Hawthorne effect, interfering with the participants responses may have occurred. Demand characteristics are an issue when the participant receives information from the training that conveys the purpose of the study and influences their responses (Huck, Cormier & Bounds, 1974). Therefore, there is a possibility that the participant will make changes in post-test responses because they believe they are supposed to make some change.

The way in which subjects are selected into the study causes a potential threat to generalizability to the extent that the entire group of subjects (both experimental and control groups) may differ from the overall population of psychiatric nurses. Nurses who chose to participate in this study may be different from the group of nurses from which they come. Since participation in this study was voluntary, the self selection of the participants presented a threat to

generalizability. This generalizability was further threatened by the lack of randomization of subjects. Group characteristics emerged that make it unlikely that a similar sample would be drawn at random from the population and therefore, the groups cannot be assumed to be a representative sample of the general population and limit the generalization of results.

The final threat to external validity, using this design, is that of reactive arrangements. Reactive arrangements occur when participants are aware that they, or their responses, are being observed for the purpose of experimentation. Since the participants were aware that they were being studied, there may have been a heightened receptiveness to the material or possibly a preoccupation with trying to determine what aspects of the process were being studied, which might have interfered with performance. Both situations threaten the generalizability of the study. Ethical considerations required that the subjects knew they were in a study, so the threat was tolerated.

Overall, the planned design has adequate internal controls while having external controls that are weak. Further weaknesses have been mentioned. The design provides the most controls possible without necessitating elaborate testing conditions, participant inconvenience or large numbers of participants. The

study was done under regular classroom conditions.

Discussion of Results

The first purpose of this study, to examine whether a structured training program affected the ability of nurses to achieve accurate results using the AIMS was tested using the first three hypotheses. The expectation was that the experimental group would achieve accurate results as compared to the NIMH scores, that they would show little variability in the range of their ratings and there would be significant improvement from pretest to post-test.

Of these expectations, only the issue of variability was realized. The experimental group rated the post-test patient very consistently. The problem with the result is that they scored the patient lower than they had on the pretest and therefore moved away from the NIMH score. There are several issues relating to the content of the training and the study design which may explain these findings.

Bostrom (1988) indicated that one advantage of the NIMH training was the opportunity for novices to compare their scores with the ratings of the expert. It is true that ample time was taken after each practice example to rate and discuss the ratings and movements seen. However, the format of how the practice patients are presented and the consistency of

the NIMH scoring may make it difficult for the novice rater to make much sense out of how NIMH came to the scores offered. Other studies (Germer et al., 1984; Lane et al., 1985; Whall et al., 1983) do not clearly address how the novice raters compared to the expert ratings.

As seen in Table 4 (page 35), examples given on the NIMH tape show the last two example patients as having overall severity (item 8) ratings of 2, yet one patient had a highest rating of 3, and ratings of at least 1 or higher in all seven body areas while the other patient had a highest rating of 2, and had only four body areas affected. With only a few examples presented and apparent inconsistency in ratings, it seems that novices to the AIMS would be confused about the scoring of the instrument.

The severity of TD demonstrated via the NIMH tape may also have affected the post-test ratings of the experimental group. The examination procedure demonstration patient as well as the first two practice example patients in the NIMH training had very severe movement disorders. Participants in the study were overheard remarking to one another about the severity of the movements seen. This researcher interpreted the remarks as an indication that the severity of movements shown on the video repulsed sone of the participants.

Clinicians who are not used to seeing patients with movement disorders may very well view a patient's movement disorder as being more severe than would a clinician who is used to seeing patients with abnormal movements and has developed a repertoire of experiences with patients which include mild to severe movement disorders.

In summary, the effect of a structured training program on the ability of nurses to achieve accurate results with the AIMS is inconclusive. In this study, the experimental group moved away from the target rating and this may have been due to the ambiguity of the AIMS rating procedures or clinical decision making cues, and the combined effect of demand characteristics coupled with a presentation which first presents a severe case followed by mild cases during the training. The experimental group did rate the post-test patient consistently and this is attributed to the training. This finding supports Lazare's (1979) assertion that the more training a group receives, the more they score alike. The control group, while there are concerns about randomization in the study, did not vary from pretest to post-test ratings, and this points to the training as the most likely cause of the low variation of experimental group scores.

As an addendum to this discussion, there is an

important clinical issue around accuracy with the NIMH ratings versus consistency of ratings within a clinical setting. For monitoring purposes, it is more important that clinicians within a specific setting are rating their patients consistently and with high inter-rater agreement or else it will be the inconsistency of the raters that will be monitored rather than the course of the patient's movement disorder. If a clinical group either overrates or underrates as a rule, the group will eventually adjust to decide when they will do further work-up for the patient (personal communication, L. M. Toenniessen, November 30, 1988).

The second question examined through this study was to what extent do the experimental group's post-test results differ from those of the control group who have read the instructions only before rating the patient? To examined this question, hypotheses two, three, and four were evaluated. As stated in the discussion above, the experimental group's rating on the post-test did vary significantly from that of the control group's mean rating, but in the opposite direction from what was expected.

Reasons for the direction of the changes in the experimental groups scoring have been proposed above. The important issue for this question is how comparable are the groups? As discussed in chapter two, it is

important to have the experimental and control groups appear as equivalent as possible so as to eliminate alternative hypotheses for the explanation of the results. In this study, the only measured significant difference between the experimental and control groups is the high amount of prior clinical use of the AIMS by the control group. In a quasi-experimental design, using a pretest, post-test, non-equivalent control group format, a final test for differences between the two groups is the comparison of the groups before the experimental treatment, or training (Polit & Hungler, 1987). Since there was no significant difference between the two groups' mean pre-test scores, it appears their ability to use the AIMS was equal at that point. Therefore, it is concluded that the training was the most likely cause of the experimental group's changes in mean score and variability.

The control group's mean post-test score was closer to the desired NIMH rating, but not significantly so. This trend may have to do with a pretest/post-test effect. It may be that with their experience at using the AIMS, the control group members were able to observe more accurately on the post-test because they had a second chance to evaluate the patient without the issues of severity and rating ambiguity in the training.

The final question examined through this study focuses on to what extent experience in psychiatric nursing affects the ability of nurses to perform the AIMS accurately. The participants happened to be nurses who have been in the field for several years and therefore, the numbers required for statistical analysis were not obtained. Although hypothesis number five called for the comparison of AIMS accuracy for nurses with less than one year of experience in psychiatric nursing, an analysis was done using the experience level set at less than five years. This revealed only indications that the hypothesis may have some support if tested on a larger scale and with novice psychiatric nurses. Benner (1984) points to five years as being a point at which nurses attain expert assessment skills, and therefore, the spread of experience in this analysis may not provide enough of a difference between the groups to be statistically significant.

Finally, there remain issues not examined through the hypotheses of this study but meriting some mention. These are issues of nurses' interest in the subject of TD monitoring, the quality of the tape, and the interest-holding aspects of the training tape. Due to the difficulty in getting nurses to participate in the study, and the sparse attendance at the training

sessions, the researcher wonders if there is a lack of interest in the subject of TD or possibly an avoidance to an iatrogenic illness.

There were several direct comments made about the quality of the tape. While its clarity was admittedly poor, the researcher felt that the essential abnormal movements were visible. However, not being able to see clearly is distracting and irritating, as evidenced by other comments during the trainings. The tape is old, but it does demonstrate mild to severe TD. To do further trainings, the researcher would use a remake of the NIMH video tape format using a new set of patients scored by clinical TD experts which would improve the technical quality and provide a mechanism to have rationale for the scoring decisions.

Limitations of the Study

As with any study, this study has limitations. The limitations exist in the areas of sampling and the ability to generalize from the study results.

The sample for the study was largely self selected. While participants were invited to attend the training sessions, no attempt to randomly draw from the population of psychiatric nurses was made. In addition, due to the necessity of offering the training as an inservice, after the initial training session where participants could be randomly assigned to

groups, all participants could not be randomly assigned. The participant characteristics may have influenced the internal validity of this study to some extent, but that remains unknown.

The external validity of the study is limited by the small number of subjects and the failure to completely randomly assign the subjects into groups. The results of the study should not be considered generalizable without further study.

Summary and Conclusions

This study has examined the research questions as stated earlier. From the analysis of the data and consideration of the results of the study, it is determined that the structured training on the use of the AIMS does make a difference over merely reading the instructions. While the training does not appear to bring the trainees toward accuracy with a normative standard score, it does appear to bring the group toward agreement within the group. This agreement probably has important clinical relevance in that it is essential that patients be rated similarly within a clinical setting to achieve effective monitoring. Also an attempt was made to examine the relationship of experience in psychiatric nursing with the ability to use the AIMS without formal training. While the study was inconclusive on this issue, the trend shown does

support the idea that general experience with psychiatric patients may enhance the raters abilities to identify subtle abnormal movements.

The findings of the study did not support any of the 5 hypotheses. The findings did point toward the training as the likely cause for the experimental group to have high agreement in the way the group rated the test patient. It may be that the training itself led the group toward a unified way of thinking about how to rate patients, but also skewed the thinking of the group to under-rate the patient.

There is a natural inclination to assume that training developed by such groups a NIMH would be appropriate and effective when in fact they may have problems not evident until explored more empirically. The results suggest that there may be some problems with the training program or the instrument itself. To further understand this, a qualitative approach may be used. Discussions with the subjects in a systematic manner to understand their responses may be most helpful in understanding these results.

Implications for Nursing Practice

Regardless of the results of the study, nurses should continue to learn about and evaluate their patients for symptoms of TD. It is not reasonable to generalize from this study that training in general is

not effective, but much more reasonably, the implication is that the subject should be further researched. The positive result of the study is that the experimental subjects rated the patient more similarly after the training. In clinical situations, it is important that clinicians can agree on their observations. Of course, it is also important that their observations be accurate as well.

A more general implication for nursing practice is that instruments are not sacred, even though they have wide clinical use. The limitations of an instrument are not known unless they are studied.

Recommendations for Further Study

The recommendations for further study of AIMS training include the replication and expansion of this study. Suggested changes involved in such a replication would be as follows:

1. Replication of this study with the following changes:
    a. Use of a larger sample
    b. Use of a new training tape of the same format as the NIMH tape, as well as the NIMH tape, except that the patients would have mild TD symptoms.
    c. Assure randomization based upon all relevant criteria.

d. Include student nurses as participants to help determine the effect of experience upon learning to use the AIMS

e. Expand the analysis to include all of the items of the AIMS.

2. Study Nurses knowledge of TD with an exploration and description of what feelings nurses have on the subject, and then explore the interaction of these feelings with the subjects' test scores.

3. A qualitative study examining the appropriateness and effectiveness of the NIMH training in relation to clinical use and clinician understanding of the AIMS items.

4. Study the relationship between the scoring of item 8 of the AIMS and the summed scores of items 1 through 7.

Bibliography

American Nurses' Association (1982). Standards of psychiatric and mental health nursing practice. Kansas City: Author.

American Psychiatric Association (1979). Tardive dyskinesia: Report of the American psychiatric association task force on late neurological effects of antipsychotic drugs. Washington, D. C.: Author.

Beck, C. M., Rawlins, R. P. & Williams, S. R. (1984). Mental health--psychiatric nursing: A holistic life-cycle approach. St. Louis: C. V. Mosby.

Benner, P. (1984). From novice to expert: Excellence and power in clinical nursing practice. Menlo Park: Addison-Wesley.

Bostrom, A. C. (1988). Assessment scales for tardive dyskinesea. Journal of Psychosocial Nursing and Mental Health Services. 26(6), 8-12.

Brown, P. & Funk, S. C. (1986). Tardive dyskinesia: Barriers to the professional recognition of an iatrogenic disease. Journal of Health and Social Behavior. 27(6), 116-132.

Burr, M. L. & Elwood, P. C. (1985). Research and development of health promotion services -- screening. In W. W. Holland, R. Detels & G. Know (Eds.). Oxford Textbook of Public Health. Vol. 3 Investigative Methods in Public Health. Oxford: Oxford University Press.

Campbell, D. T. & Stanley, J. C. (1966). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Chien, C., Jung, K, & Ross-Townsend, A. (1980). Methodological approach to the measurement of tardive dyskinesia: Piezoelectric redording and concurrent validity test of five clinical rating scales. In W. E. Fann, R. C. Smith, J. M. Davis, and E. F. Domino (Eds.). Tardive dyskinesia: Research and treatment (pp. 233-241). Jamaica, NY: Spectrum.

Chu, G. C. & Schramm, W. (1967). Learning from television: What the research says. Educational Resources Information Center Report No. ED-014 900. Stanford University, Stanford.

Crane, G. E. & Narango, E. R. (1971). Motor disorders induced by neuroleptics. Archives of General Psychiatry. 24(2), 179-184.

Crawford, A. L. & Kilander, V. C. (1985). Psychiatric and mental health nursing (6th ed.). Philadelphia: F. A. Davis.

Critchely, D. L. & Maurin, J. T. (1985). The clinical specialist in psychiatric mental health nursing. New York: John Niley & Sons.

DeVaugh-Geiss, J. (1982). Tardive dyskinesia: Phenomenology, pathophysiology and pharmocology. In J. DeVaugh-Geiss (Ed.), Tardive dyskinesia and related involuntary movement disorders: The long-term effects of antipsychotic drugs. Boston: John Wright.

Flaskerud, J. H. & Van Servellen, G. M. (1985). Community mental health nursing: Theories and methods. Norwalk: Appelton-Crofts.

Germer, C. K., Seraydarian, L., McBrearty, J. F. (1984). Training hospital clinicians to diagnose tardive dyskinesia. Hospital and Community Psychiatry. 38(8), 769-783.

Gibbons, J. F., Kincheloe, W. R., & Down, K. S. (1977). Tutored videotape instruction: A new use of electronics media in education. Science. 195(4283), 1139-1146.

Gualtiere, C. T. & Sprague, R. L. (1984). Preventing tardive dyskinesia and preventing tardive dyskinesia litigation. Psychopharmacology Bulletin. 20(3), 346-348.

Guy, W (Ed.) (1976). EDCUE Assessment manual for psychopharmacology. Rockville: U.S. Department of Health, Education, and Welfare.

Haber, J. H., Leach, A. M., Schudy, S. M., & Sideleau, B. F. (1982). Comphrehensive psychiatric nursing (2nd Ed.). New York: McGraw-Hill.

Hagerty, B. K. (1984). Psychiatric--mental health assessment. St. Louis: C. V. Mosby

Huck, S. W., Cormier, W. H., and Bounds, W. G. (1974). Reading statistics in research. New York: Harper & Row.

Jeste, D. V. & Wyatt, R. J. (1982). Understanding and treating tardive dyskinesia. New York: Guilford Press.

Kane, J. M., Woener, M., Weinhold, P., Kinon, B., Leiberman, J., and Wegner, J. (1983). Epidemiology of tardive dyskinesia. Clinical Neuropharmacology. 6(2), 109-115

Kazamatsuri, H., Chien, C. P., & Cole, J. O. (1972). Treatment of tardive dyskinesia. III. Archives of General Psychiatry. 27(12), 824-827.

Lane, R. D., Glazer, W. M., Hansen, T. E., Berman, W. H., & Kramer, S. I. (1985). Assessment of tardive dyskinesia using the abnormal involuntary movement scale. The Journal of Nervous and Mental Disease. 173(6), 353-357.

Lazare, A. (1979). Outpatient psychiatry: Diagnosis and treatment. Baltimore: Williams and Wilkins.

Loney, J. (1980). Hyperkinesis comes of age: What do we know and where should we go? American Journal of Orthopsychiatry. 50(1), 28-42.

Munetz, M. R. (1985). Overcoming resistance to talking to patients about tardive dyskinesia. Hospital and Community Psychiatry, 36(3), 283-287.

Polit, D. F. & Hungler, B. P. (1987). Nursing research: Principles and methods (3rd ed.). Philadelphia: J. P. Lippincott.

Richardson, M. A. & Craig, T. J. (1982). Tardive dyskinesia: Inter- and intra-rating scale comparisons. Psychopharmacology Bulletin. 18(1), 4-6.

Simpson, G., Zoubok, B. & Lee, J. H. (1976). An early clinical and toxicity trial of EX11-582A in chronic schizophrenia. Current Therapy Research. 19(6), 87-93.

Sox, H. C., Jr., Marton, K. I., Higgins, M. C., & Hickman, D. H. (1984). Tutored videotape-instruction in clinical decision-making. Journal of Medical Education. 59(3), 188-195.

Sprague, R. L., White, D. M., Ullman, R., & Kalachnik, J. E. (1984). Methods for selecting items in a tardive dyskinesia rating scale. Psychopharmacology Bulletin. 20(3), 339-345.

Stuart, G. W. & Sundeen, S. J. (1983). Principles and practice of psychiatric nursing (2nd ed.). St. Louis: C. V. Mosby.

The Task Force on Late Neurological Effects of Antipsychotic Drugs. (1980). Tardive dyskinesia: Summary of a task force report of the American Psychiatric Association. American Journal of Psychiatry. 137(10), 1163-1172.

Whall, A. L., Engle, V., Edwards, A., Bobel, L., & Haberland, C. (1983). Development of a screening program for tardive dyskinesia: Feasibility issues. Nursing Research, 32(3), 151-156.

Wilson, H. S. & Kneisl, C. R. (1983). Psychiatric nursing. Menlo Park: Addison-Wesley.

APPENDIX A

EXAMINATION PROCEDURE

Either before or after completing the Examination Procedure observe the patient unobtrusively, at rest (e.g., in waiting room).

The chair to be used in this examination should be a hard, firm one without arms.

---

1. Ask patient whether there is anything in his/her mouth (i.e., gum, candy, etc.) and if there is, to remove it.

2. Ask patient about the <u>current</u> condition of his/her teeth. Ask patient if he/she wears dentures. Do teeth or dentures bother patient now?

3. Ask patient whether he/she notices any movements in mouth, face, hands, or feet. If yes, ask to describe and to what extent the <u>currently</u> bother patient or interfere in with his/her activities.

4. Have patient sit in chair with hands on knees, legs slightly apart, and feet flat on floor. (Look at entire body for movements while in this position).

5. Ask patient to sit with hands hanging unsupported. If male, between legs, if female and wearing a dress, hanging over knees. (Observe hands and other body areas.)

6. Ask patient to open mouth. (Observe tongue at rest within mouth.) Do this twice.

7. Ask patient to protrude tongue. (Observe abnormalities of tongue movement.) Do this twice.

* 8. Ask patient to tap thumb, with each finger, as rapidly as possible for 10-15 seconds; separately with right hand, then with left hand. (Observe facial and leg movements.)

9. Flex and extend patient's left and right arms (one at a time.)

10. Ask patient to stand up. (Observe in profile. Observe all body areas again, hips included.)

* 11. Ask patient to extend both arms outstretched in front with palms down. (Observe trunk, legs, and mouth.)

* 12. Have patient walk a few paces, turn, and walk back to chair. (Observe hands and gait.) Do this twice.

* Activated Movements

Abnormal Involuntary Movement Scale
(AIMS)

Instrutions: Complete the examination procedure before making ratings.
MOVEMENT RATINGS: Rate highest severity observed. Rate movements that occur upon activation one less than those observed spontaneousley.

Code: 0 = None 1 = Minimal 2 = Mild 3 = Moderate 4 = Severe

1. **Muscles of Facial Expression** e.g., movements of forehead, eyebrows, periorbital area, cheeks; including frowning, blinking, smiling, grimacing	0	1	2	3	4	
2. **Lips and Perioral Area** e.g., puckering, pouting, smacking	0	1	2	3	4	
3. **Jaw** e.g., biting, chewing, clenching, mouth opening, lateral movement	0	1	2	3	4	
4. **Tongue** Rate only increase in movement both in and out of mouth, NOT inability to sustain movement.	0	1	2	3	4	
5. **Upper Extremities** (arms, wrists, hands, fingers) Include choreic movements, (i.e., rapid, objectively purposeless, irregular, spontaneous), athetoid movements (i.e., slow, irregular, complex, serpentine). DO NOT include tremor (i.e., regular, repetitive, rhythmic)	0	1	2	3	4	
6. **Lower Extremities** (legs, knees, ankles, toes) e.g., lateral knee movement, foot tapping, heel dropping, foot squirming, inversion and eversion of foot	0	1	2	3	4	
7. **Neck, shoulders, hips** e.g., rocking, twisting, squirming, pelvic gyrations	0	1	2	3	4	
8. **Severety of abnormal movements**	0	1	2	3	4	‡
9. **Incapacitation due to abnormal movements**	0	1	2	3	4	
10. **Patient's awareness of abnormal movements** Rate only patient's report	0	1	2	3	4	
11. **Current problems with teeth and/or dentures**	No Yes		0 1			
12. **Does patient usually wear dentures?**	No Yes		0 1			

‡ Item used for comparison with NIMH standards.

APPENDIX B

Demographics Questionnaire

Please answer the items in the left column.

Coding

Current age _____years _____

Female_____ Male _____ _____

How long have you been working as an R.N.?
_____yrs _____mos _____

Highest level of nursing education
AD____ Diploma____ BS____ MN____
PhD___ _____

How long have you been practicing
psychiatric nursing? _____yrs_____mos _____

How many years in each type of psychiatric nursing?
Inpatient _____ _____
Outpatient _____ _____
Private Practice_____ _____
Educational _____ _____

Have you ever used the AIMS before in
clinical practice?
Yes____ No____ _____
If yes, how many times?
1-10 ___
11-20 ___
21-30 ___
31+ ____ _____

Have you ever had formal AIMS training?
Yes____ No____ _____
If yes, please briefly explain._____
_____
_____ _____

APPENDIX C

Oregon Health Sciences University

School of Nursing

CONSENT FORM

"Using the Abnormal Involuntary Movement Scale (AIMS) for Tardive Dyskinesia Screening" by Davis C. Clowers, R.N., P.M.H.N.P., B.S.N. Under the supervision of Sarah Porter-Tibbetts, R.N., M.P.H., M.S.

My participation in this experimental study will take approximately three to three and one half hours, during which time I will fill out AIMS evaluations and observe a training videotape on the subject of evaluating tardive dyskinesia (TD) using the AIMS.

The potential benefits of this study are that the participants will gain an increased understanding of how to screen patients for TD and in general, the patients of the participants will benefit through more expert monitoring for the signs of TD.

There are no foreseeable risks involved in my participation in this study. All information obtained will be kept confidential. Code numbers will be given to each person to protect his or her privacy. Identifying information will be destroyed after the project is completed. Neither my name nor my identity will be used for publication or publicity purposes.

Davis Clowers has offered to answer any questions regarding this study and may also be reached at 228-7134 or 282-1581 if questions arise.

The Oregon Health Sciences University, as an agency of the State, is covered by the State Liability Fund. If you suffer any injury from the research project, compensation would be available to you only if you establish that the injury occurred through the fault of the University, its officers or employees. If you have further questions, please call Dr. Michael Baird at (503) 279-8014.

I understand I may refuse to participate or withdraw from this study at any time without affecting my relationship with, or treatment at, the Oregon Health Sciences University.

I have read the foregoing and agree to participate in this study.

Date __/___/___ Signature of Subject _____

APPENDIX D

Sample Letter

Dan Kamada, RN, Head Nurse
Psychiatric Crisis Unit
OHSU, Hospital North
3181 SW Sam Jackson Park Road
Portland, OR 97201

Dear Dan,

Attached is an announcement of a free training offered for psychiatric nurses. The subject is the use of the Abnormal Involuntary Movement Scale (AIMS). The AIMS is a screening instrument for detecting abnormal movements resulting from the use of neuroleptic medication.

I would appreciate your posting of the anouncement. I also ask that you encourage your staff to attend if possible. The training should take no longer than 3 1/2 to 4 hours, and your staff will learn a valuable assessment procedure.

I would like to point out that the training is part of my Master's research project and some of the imformation gathered from the participants will be used as part of that project. Of course, each participant's anonymity will be protected and participants will not be at any risk. The announcement also states that the training is part of a research project.

If you have any questions about this matter, you may contact me at home (282-1581) or at work (228-7134).

Sincerely,

Davis C. Clowers RN, PMHNP
3236 NE 18th Ave.
Portland, OR 97212

APPENDIX E

Training Announcement

You are invited to attend a

Training

H O W T O U S E T H E A I M S
(Abnormal Involuntary Movement Scale)

Date:
Place:
Time:

Purpose: This training is being offered as part of a nursing Master's research project to train nurses in the use of the Abnormal Involuntary Movement Scale (AIMS). Participants will not be at any risk. The information offered at the training will help participants develop skills in assessing abnormal movement disorders such as tardive dyskinesia.

Time Involved: The training should take no more than four hours.

Cost: Free

Eligability: RN's practicing in settings with psychiatric patients.

Trainer: Davis Clowers is a PMHNP employed at Mental Health Services West in Portland. He is also a Master's candidate at the OHSU School of Nursing. He has been using the AIMS in clinical practice for eight years.

Please call Davis Clowers at (503) 282-1581, evenings, or (503) 228-7134, days, to pre-register by date. Space may be limited.

Appendix F

Schedule of Events
Experimental and Control Groups

	:Start	:15	:30	:45	1:00	1:15	1:30	1:45	2:00	2:15	2:30	2:45	3:00	3:15	3:30
Exp.	Intro. Consents Pretest	NIMH Training				Break				Break		Post-Test	TD Lecture		
	:----------Data Collection Period----------:														
Control	Intro. Consents Pretest	TD Lecture		Post-test		Break	NIMH Training								end
	:--------Data Collection Period ---------:														

AN ABSTRACT OF THE THESIS OF

Davis C. Clowers

For the MASTER OF NURSING

Date of Receiving this Degree: June 9, 1989

Title: THE EFFECTS OF TRAINING ON NURSES' PERFORMANCE USING THE ABNORMAL INVOLUNTARY MOVEMENT SCALE.

Approved: [redacted]

Sarah Porter-Tibbetts R.N., M.P.H., M.S.,

Thesis Advisor

The nursing literature encourages nurses to evaluate their patients for the signs of tardive dyskinesia (TD). However, training programs in the techniques of TD assessment are scarce. Nurses have been using assessment instruments, such as the Abnormal Involuntary Movement Scale (AIMS) for years without training. This experiment was designed to evaluate the performance of nurses using the AIMS with and without training. Psychiatric nurses attended a training in how to use the AIMS. Attempts were made to randomly assign the participants to an experimental group and a control group. Their pre- and post-test scores on a global item of the AIMS were used to evaluate their performance using the AIMS as compared to the standard scores of the NIMH training videotape that was presented.