The University of Texas at Arlington McNair Research Journal

The University of Texas at Arlington McNair Research Journal

Summer 2000 • Volume 4

The University of Texas at Arlington
McNair Research Journal
Summer 2000 • Volume 4

Table of Contents

Notes from the Director

It is with much pride and respect that I congratulate the Summer 2000 McNair Scholars at The University of Texas at Arlington on their outstanding research efforts. As their work suggests on the following pages, they have shown intense commitment to the pursuit of new knowledge in their respective fields. Their scholarly achievements deserve recognition and merit praise from their peers.

During the first year of the new millennium, it seems appropriate to showcase the high caliber of research which can be accomplished by talented undergraduate students through the guidance and direction of interested faculty mentors and with the encouragement and support of program staff. The following papers represent eleven weeks of research accomplished jointly by scholar and mentor. Each scholar wrote and presented his or her paper at a formal campus event, and frequently at subsequent undergraduate research conferences. We hope the growth and academic development of our scholars through the research experience has better prepared each to continue pursuing academic excellence.

For their ongoing support and assistance, we thank Dr. Dale A. Anderson, Vice President and Dean of Research and Graduate Studies; Dr. Mary R. Ridgway, Vice President for Undergraduate Academic and Student Affairs; and each of our faculty mentors who gave unselfishly of their time, talent, and knowledge.

Best wishes, McNair Scholars, as you continue your academic journey in this new century!

Kathryn Head

Acknowledgements

McNair Scholars Program Staff

McNair Scholars Program Director .... Kathryn Head
Assistant Director .... Dr. Joan Reinhardt
Secretary .... Cassie Davis
Editorial Assistant .... Theresa Alvarado-Cunningham

Faculty Mentors

Dr. Cheryl Anderson, School of Nursing
Dr. Dana Austin, Department of Sociology and Anthropology
Dr. Mark Burleson, Department of Biology
Dr. Perry N. Fuchs, Department of Psychology
Dr Thomas Gluick, Department of Chemistry and Biochemistry
Dr. Kendall Harris, Department of Mechanical Engineering
Dr. D.L. Hawkins, Department of Mathematics
Dr. Sheik N. Imrhan, Department of Industrial Engineering
Dr. John M. Kennedy, Department of Mechanical Engineering, Clemson University
Dr. Peter Lehmann, School of Social Work
Dr. Frank K. Lu, Department of Mechanical and Aerospace Engineering
Darren McHenry, Department of Music
Dr. Charles Nussbaum, Department of Philosophy
Dr. Arthur A. Reyes, Department of Computer Science and Engineering
Dr. Joan R. Rycraft, School of Social Work
Dr. Marilee Schmelzer, RN, School of Nursing
Dr. Pat Taylor, School of Architecture

The University of Texas at Arlington
McNair Research Journal Authors

María José Brown,
Industrial Engineering

José R. Espino,
Computer Science
Engineering

Jennifer E. FitzGerald,
Nursing

DeAdera Gardner,
Philosophy

Rachel Hock,
Interdisciplinary
Studies

Tina Holder,
Biochemistry

David S. Jensen,
Aerospace Engineering
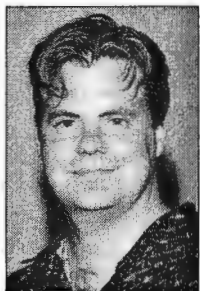

Terrence E. Johnson,
Mechanical Engineering

The University of Texas at Arlington
McNair Research Journal Authors

Damien Kinzler,
Psychology

Shirley Mathews,
Architecture

Tamara Dawn Robertson,
Anthropology

Joseph Rodriguez,
Music

Allyson J. Schuerman,
Nursing

Philip Silva,
Biology

Ni Khoi Tri,
Mathematics

Julie B. Williams,
Nursing

Place Attachment vs. Place Alienation:
The Impact of Design on Student Satisfaction with a
Metropolitan University Campus Environment

Shirley Mathews
Mentor: Dr. Pat Taylor, School of Architecture

ABSTRACT

This research explored 'sense of place' as related to image, and its effect on student attachment to The University of Texas at Arlington, located in the Dallas-Fort Worth metroplex. It also studied students' perceptions of campus image and how they linked to 'attachment to place.' Place image was examined by method of student-drawn maps. Interviews elicited responses on spirit and sense of place and related meanings. Problems with the physical design (lack of places to socialize, lack of a pedestrian-friendly campus, and perceptions that the campus is specifically designed as a "commuter school") were shown to affect the behavior of students in their use of campus. Many students described the placelessness of and feelings of disenfranchisement with their campus. A sense of placelessness on campus was shown to be related to students' limited place attachment.

INTRODUCTION

The 'sense of place' associated with a university is very important to the image students have of their campus. It is what graduates remember decades after they have left, often prompting alumni to contribute money to the university, and it is what visiting high school seniors remember when they decide which university to attend. A sense of place is made up of physical attributes, meanings ascribed to a place and resulting feelings about a place. Theorists have defined the nuances of meanings about places and people's interaction with a place in an attempt to understand what can make one place better than another. Literature on 'place' defines the different aspects of 'place' this way:

> Place has two aspects: The sense of place, which is the particular experience of a person in a particular setting, (feeling stimulated, excited, joyous, expansive, and so forth); and the spirit of place, which is the combination of characteristics that give some locations a special "feel" or personality (such as a spirit of mystery or of identity with a person or group) (Steel 1981, p. 11).

Research has been done to examine why people experience places in different ways and how people interact with their environments. Some places, such as cities, neighborhoods or districts, elicit more favorable feelings than do other places. Landscape architect and city planner K. Lynch explains:

> We have powerful abilities for recognizing places and for integrating them into mental images but the sensory form of those places can make that more or less difficult. So we take delight in physically distinctive, recognizable locales and attach our feelings and meanings to them. They make us feel at home, grounded. Place character is often recalled with affection; its lack is a frequent subject of popular complaint . . . indeed, a strong sense of place supports our personal identity (Lynch 1976, pp. 24-25).

These physically distinctive places are more memorable and have more clarity in the minds of the users. People create mental maps of a place to help locate themselves in their surroundings. Research has shown that drawn maps reflect the image of a place as perceived by the sketcher of the map (Lynch 1960). Clear mental images of a place relate to the popularity of a place.

In contrast to environments that possess distinctive and desirable characters are those environments that are "placeless." A placeless environment not only does not possess those memorable and distinctive features but can also foster a sense of place alienation. The concept of placelessness has been associated with a variety of negative perceptual and affective responses to the environment that can result from increased homogeneity of the built environment, disruption in one's sense of continuity with a place, and other factors (Relph 1976; Garnham 1985; Low & Altman 1992; Giuliani & Feldman 1993).

The relationship between landscape preferences and landscape content has been studied from the perspective of cognitive or psychological methodological approaches, which seek to relate values of preference and satisfaction to different types of environmental landscapes, including cities, neighborhoods, and countryside (Kaplan 1982, 1985). Qualities of a residential neighborhood's image relate strongly to the place attachment residents felt (Hull 1992). A link between an individual's attachment to place and the relationship between community conceptions of general environmental quality and associated physical features has also been studied using criteria of image congruity:

> Image congruity refers to the fit between the meanings and values associated with a place and a person's image of self. The landscaping, layout, and personalization of a suburban house may communicate one's self image to others. It informs neighbors and visitors something of the owner's values, beliefs, life-stage and life-style (Hull 1992, p.181-191).

Place attachment promotes social interaction and a sense of community.

No empirical research has been directed at determining what may be involved in a student community's perception of campus image or sense of place and how this affects place attachment. This research will examine students' perceptions of campus image and how this links to attachment to place.

METHOD

This is a qualitative study of the experiences of students attending The University of Texas at Arlington (U.T. Arlington). Located in Arlington, Texas, between Dallas and Fort Worth, U.T. Arlington is a major public general academic university. Enrollment in the Fall 1998 semester was 18,662. U.T. Arlington primarily draws on the population of the greater metropolitan area for enrollment.

U.T. Arlington is characterized by a general lack of school spirit and the small number of students who spend time on campus. The attitudes and behaviors of students were studied using the interview as the main instrument of research. In this case, "questions designed to measure the variables of the research were included in the interview schedule. These questions were then considered as items in a measurement instrument, rather than as mere information gathering devices" (Kerlinger 1986, p. 440). No formal hypotheses were formulated; however, a set of research questions was developed to help explore possible relationships between landscape features and meanings involved in the student community's perceptions of the university campus.

In order to examine the sense of place on the U.T. Arlington campus both the image, or features of the space, and the conceptions and meanings that students associate with the campus setting were examined (Canter 1977). The method of using drawn maps to ascertain the image of the campus in the minds of students was also used (Lynch, 1960). Twenty

personal interviews of U.T. Arlington students were conducted by the researcher, using a tape recorder and notebook. Respondents ranged evenly in age from 19 to 48 years old. Ten students were living on campus or within a 15-minute drive of campus, and the remaining ten students lived 15 to 90 minutes away from campus. Every school or college on campus was represented.

Phase I interviews consisted of three parts: (a) an essay type question, (b) specific questions soliciting directions to an across campus destination, and (c) a drawing of the main parts of campus. The essay question, "What do you think of the U.T. Arlington campus?" was intentionally vague to prompt students to draw their own meanings from the question. The map-drawings along with students' reactions (frustration, confusion, joy), as they described giving directions to an across-campus location, constituted raw data for later interpretation.

Phase II interviews continued use of the maps, as it was shown to help focus responses on the physical experience of the campus. Interview questions were expanded into a 12-question interview designed to elicit responses on spirit and sense of place, use, and meanings associated with the U.T. Arlington campus.

RESULTS

The physical layout of campus as drawn by interviewees shows academic buildings floating in an undifferentiated plane. Connections between buildings by paths are rarely noted. The elements most important to students were Cooper Street, the fountain at the central bridge, and the main library and its mall to the west. These elements were drawn first or larger, and other elements were drawn as related by the proximity to these elements. Relationships between buildings and landmarks could be ascertained from the maps.

When asked to draw an overview of the U.T. Arlington (UTA) campus, most students drew only the pedestrian core of campus. Elements such as the Activities Building, Greek row, and the stadium were not drawn on the maps. Students voiced that these places were near, but not on campus. The lack of character in the U.T. Arlington campus is apparent from the almost bare maps. Distinctive elements were very hard to identify. Building-to-building connections and relationships were rarely shown and were difficult for students to describe. In studying the individual sketches and interviews, it became apparent that very few of the respondents had a comprehensive view of the university at which they spent so much time. As one student noted:

It's not attractive, not cohesive. It was very hard to get around at first. I only walked the perimeter of campus, never across, because I couldn't tell where I was. The tuition is reasonable and the programs can be good, but poor little UTA needs some identity. You know, character. There is a real lack of design here.

Most students did not feel comfortable walking across campus and primarily accessed different parts of campus from the peripheral parking lots. Poor orientation and students' inability to describe or differentiate the parts of campus became clear when they were asked to give directions to an across-campus destination, as three students revealed below:

That's the hardest thing about UTA, is trying to give somebody directions. There are no physical landmarks. You can't say "The statue of so-n-so", or "That thing on the right" or "If you've gotten to this you've gone too far." When I try to tell people where Preston Hall is, that's a hard one.

***

Very hard to navigate campus without knowing what all the buildings are. You have to navigate along the edges of buildings, because there are no public spaces. I'll walk past that van with the American flag painted on it. That van [the *Shorthorn*'s van] is always in the same spot. If you're going the right way you'll go right past it.

***

I still think you ought to drive it, because it's easier. Just go down Cooper to Border and straight to Pecan. It is very hard to tell someone how to get around by walking. You just get used to it. I would just have to take them there, I guess.

Interviewing proved to be a difficult task. Articulating what they could say about the University campus was a problem for students. When asked for a general characterization of the campus, students often noted other places or other universities for which they could answer the question, making it clear that U.T. Arlington did not possess those attributes. Campus is at best 'nice' as another student remarked:

I think we have a nice campus. I mean, it's not as traditionally styled as some campuses. It seems like it has more modern looking buildings as opposed to, say, TCU. It's arranged differently–the buildings and everything are -- the way they are designed are -- uh, they're just designed different. UTA isn't a young school by any means, but the buildings look modern. It would be nice if the buildings were nicer, but it's not necessary. It's a nice campus; they keep it up nice and keep the grounds well.

Many students indicated the lack of a "dynamic" campus center or core. The open space to the west of the library is the most important and meaningful on campus as a recognizable public space. Students appreciated the open view across the space, which gives clarity and recognizability to the place. Students felt this area should provide the most services and amenities; many commented that the university center should be located next to the library mall because of its importance on campus, as several students indicated:

People may quit complaining so much about the parking situation if there were things to look at, and places that were enjoyable; conversation places.

***

It seems like maybe because it's UTA they say "Well, it's just UTA, and so we can just do this much for UTA." We're the same level as UT or anybody else. We're a university campus and we should enjoy the same perks and enjoyments as any university campus. There are no places to stick, to find yourself at, to congeal at.

***

Something like university center should be smack in the middle. Even though it's a poor student union if it was better more people would...I think it should be in the center. Other campuses have a UC that is the main focus. But I don't go there, and can't tell anyone how to get there! Not much goes on there anyway, I don't go there to hang out.

Town squares of European cities and the village greens of East Coast towns are the traditional places where the public gathers. They unite a city, town, or village (Whyte, 1980), allowing people to eat, talk, stroll, or rest together. Our university campus has only one place with this potential. Student comments show there is no place truly functioning this

way. Many people asked for well-designed places to sit that would foster conversation. Most students asked for a simple place to sit, and most of all, a "Place of our own!" Several students express such a desire:

When I was taking the basic core stuff, and I had a few English professors who would very often say, "If you ever want to discuss some of this, I'd be more than happy to go to a coffee shop with you, and sit down and discuss some of this with you." The problem that this created, is that you have to drive across town to find a coffee shop. You have to go to Lincoln Square. There isn't anything near, that you can just walk across to after class. You can't say, "meet me at the . . ." You'd have to say, "Would you like to drive over there and meet me in 45 minutes?" By the time you get there you'd forget what you're gonna talk about! It would be such a better atmosphere, if you could say, lets walk over here. Rather than saying, lets drive, because it makes you feel...you better have something good to say.

***

The architecture courtyard is perfect for a quiet, reflective chat. But there is no dynamic. We need places that attractive and with good conversation seating in places on campus where there is a lot of activity. Then if you were hanging out some other people you know would happen by.

***

If there were a coffee shop on campus I'd spend more time on campus. If there were one near the library I'd go all the time. Campus needs something, especially something to do. A pub! Now, I would be there! We need some hangout spots where you can be boisterous. Lounges are nice but you have to be quiet.

Memorable landmarks are very few. The most frequent comment about a distinctive campus element was not a place but a street. Cooper Street is a campus landmark. It also divides the university campus into twin campuses. The lack of specific landscape details results in bland, ambiguous spaces that are not stimulating to walk past. Many students noted the lack of historical buildings on campus as one thing they missed from other campuses, as noted below by several respondents:

I think it's a problem that you can't tell when you're on the UTA campus. I think all the curbs should be painted blue. Something for an identity. I think we should be able to paint the light poles white and blue. All around campus to let people know they're on campus. Get out the school spirit, let people know what the school colors are.

***

Even the old bookstore was at least by something -- Architecture Building & Fine Arts. We knew where it was- it was by the Corral [restaurant], stuff like that. Now it's like, oh, it's way at the other end of campus by itself, you know? I wish they would have put it in a better location.

***

The campus needs to be more user-friendly. And pretty, and some places where you felt some type of feeling. It's elusive, but someplace with character, someplace interesting. That's why it's so easy to walk on campus and go "oh, wow, I'm at UTA." instead of "I've arrived at UTA!"

Changing the location of the bookstore has not been welcome. Students have called it a good idea in a bad location. They emphatically miss the walk through the architecture courtyard to the old bookstore. They see the new location as off campus and designed to be approached by car. The walking experience is unpleasant and the main entrance and landscaping faces out away from campus forcing the pedestrian to walk around to the "far side."

Although the campus is trimmed and mowed neatly, it is lacking lush, shady, well-designed areas. Any grassy spaces are cherished, showing how much they are needed, as noted below by two students:

> I like the landscaping, especially the fountain at the center bridge that runs over Cooper. Over at the architecture courtyard, it's so open. I don't think we have enough of that. Like, there is a lot of seating, places to relax over there and over here there's nothin'! I mean there's a lawn but there's nowhere to sit, no shade–you know what I'm saying? Other than that I think they keep it up pretty nice.
>
> ***
>
> I want to feel one with nature. I want to see trees and beautiful landscaping. The trees we have are nice but we need flowers, something better than just grass.

Thoughtfully designed places are mentioned by students across all majors: The Architecture Building courtyard and Johnson Creek along Mitchell Street. The fountain at the center bridge was mentioned often as being a very beautiful fountain, and the Nedderman Hall landscaping was notable, but these were seen as only places to look at, not as places to stop and relax. Two students remarked on this:

> Over by the Architecture Building where there is the fountain and its really green, you can sit down and talk, and there is a reason to stay on campus.
>
> ***
>
> Lincoln Square fountain is where people actually go and sit around a fountain. It's inviting. You don't see anybody sitting around these fountains here, because the chairs that they do put next to the fountains, a lot of the times are stuck back into the bushes. That fountain over by UH, the trees and benches that are over there are up against the wall, and the overgrown hedges will push you off the seat.

U.T. Arlington is designed for the car--campus is easy to access and easy to leave. Walking is rejected in favor of driving, as several survey participants confirmed:

> It's built for its function which is a commuter school as opposed to other schools where people live on campus. Here a large number of people can come and go fairly easily. There's a lot of parking around, whereas at a school where people live, there's not as much parking. Here it's pretty accessible to coming and going easy.
>
> ***
>
> What every high school counselor will tell you is "Go visit the campus. Ask yourself if you feel comfortable there. Is it inviting? Do you want to spent time there?" I could sum it up like this: UTA is repelling. I think that all of the departments on campus are nothing more than a glorified junior college in the sense that you come here to classes and then you go.
>
> ***
>
> When I did my core stuff, and I was taking classes on the other side of campus, there would be a dramatic shift in age after 3:00 p.m. All of the younger people go to morning

classes. The later classes, after 3:00 p.m., or at 5:30 p.m., [students] would be dramatically older, 40-years or so. Those older commuters come here mostly at night. The majority of the people at UTA are here during the day, actually walking around. And UTA doesn't address them, those younger students that want a college experience.

<div align="center">***</div>

I guess I'm classified as a commuter, but I don't feel like one. I live so close, I spend a lot of time here. The residence halls and most of the apartments are just not nice. I wouldn't want to live there.

DISCUSSION

There is a substantial daytime population of young students with nowhere to go at U.T. Arlington. Students living on campus left campus just as often and had the same feelings for campus as those students who left immediately after class because of full-time work and family responsibilities. Students who lived off campus often wanted to stay on campus for several hours until their next class, but left because the spaces were unpleasant or not what they needed. Their perception was that they were leaving because they lived off campus, when in fact they had already expressed their dissatisfaction with a lack of comfortable campus places and amenities. Leaving campus may be a rejection of the lack of design quality.

Students involved in campus activities had much higher feelings of satisfaction with the campus and U.T. Arlington. It ceases to be impersonal and isolated and becomes a warm, friendly community. These feelings transfer onto the view of the physical design only so far, however, and these students also recognize the need for a better campus. The education available at U.T. Arlington is seen as high quality, but the atmosphere on campus is less than satisfactory.

U.T. Arlington is designed for the car--campus is easy to access and easy to leave. Walking is rejected in favor of driving. Cooper Street gives the University of Texas at Arlington campus the appearance of a place to pass through rather than to stop. Most landscaped areas on the east side of campus (along Nedderman Street and near Pecan Street) are islands of greenery and color isolated by pavement. The areas are beautiful, but students felt they welcomed only those students that access the campus by car. Students describe feeling unwelcome and uncomfortable while on campus.

The University of Texas at Arlington is described as having an "urban feeling." Many tall buildings, parking lots, and a lack of greenery, color and public squares result in a campus that students use like an office park. The "urban feel" contradicts the pedestrian aspect that students associate with the idea of a university campus.

Students do not feel that The University of Texas at Arlington has an image that they relate to. This image incompatibility is shown in: 1) students seeking to alter the social image of the campus setting by personalizing their surroundings through curb painting and light pole markings; 2) students labeling themselves as commuters, when they are not; 3) students feeling disenfranchised from the campus and having limited place attachment.

CONCLUSION

The 600 new residents who will reside in Arlington Hall this fall will use the campus in much the same way as the other on-campus students do. They will use what is close to

them: the university center, the new book store, and the recreation rooms and volleyball courts designed as part of the Arlington Hall complex. However, their presence will change the campus atmosphere very little if the pedestrian core of campus is not extended and made more "user-friendly" soon.

Students described The University of Texas at Arlington as a placeless place in which they were not comfortable and had limited feelings of warmth towards the campus. Sense of place and image congruity are shown to relate to attachment to place. Architecture, landscape architecture and the planning and design professions can make a great deal of difference in the feelings of attachment to place that students have.

REFERENCES

Canter, D. *The Psychology of Place*. London: The Architectural Press Ltd., 1977.

Henderson, Karla A. *Dimensions of Choice: A Qualitative Approach to Recreation, Parks, and Leisure Research*. State College, PA: Venture Publishing, Inc., 1991.

Hull, R. Bruce IV. "Image Congruity, Place Attachment and Community Design." *The Journal of Architectural and Planning Research* 9 (Autumn 1992): 1-15.

Garnham, H. L. *Maintaining the Spirit of Place: A Process for Preservation of Town Character*. Mesa, AZ: PDA Publishers, 1985.

Guiliani, M. V. & Feldman, R. "Place Attachment in a Developmental and Cultural Context." *Journal of Environmental Psychology* 13 (1993): 267-272.

Kaplan, R. "Nature at the Doorstep: Residential Satisfaction and the Nearby Environment." *The Journal of Architectural and Planning Research* 12 (1985): 161-176.

Kaplan, S. & Kaplan, R. *Humanscape*. Ann Arbor, MI: Ulrich's Books, Inc., 1982.

Kerlinger, F. N. *Foundations of Behavioral Research, 3rd ed*. Fort Worth, TX: Holt, Rinehart and Winston, Inc., 1986.

Low, S. M. & Altman, I. "Place Attachment: A Conceptual Inquiry." In I. Altman & S. M. Low, (Eds), *Place Attachment*, Vol.12. New York: Plenum Press, 1992.

Lynch, Kevin. *The Image of the City*. Cambridge, MA: MIT Press, 1960.

Miller, P.A. "A Profession in Peril?" *Landscape Architecture* 87 (August 1997) : 66-88.

Relph, Edward. *Place and Placelessness*. London: Pion Limited, 1976.

Rodiek, J.E. "Editorial." *Landscape and Urban Planning* 32 (1995): 1-2.

Steel, Fritz. *The Sense of Place*. New York: CBI Publishing Company, Inc., 1981.

An Investigation of Finger Torque Strength

María José Brown
Mentor: Dr. Sheik N. Imrhan,
Department of Industrial Engineering

ABSTRACT

This paper describes a study to investigate finger torque strengths on individuals in an assembly line. Handles with three types of surfaces (smooth, knurled and flanged) and seven diameters were used. Both tightening and loosening strengths were tested. The experiment was performed with 23 subjects working on different schedules. Subjects were females and males between the ages of 25 and 55. No subject had earlier hand or finger injuries. The results indicated that finger torque strength depended on the type of surface used, the direction of the torque and diameter of the handle. Knurled handle torque strength was stronger than smooth handle strength, tightening greater than loosening, and strength decreased as handle size decreased. Ergonomically, tool design should consider tool surface as well as finger holding size. Further studies may be performed to consider finger strength with different finger positions.

1. INTRODUCTION

Many studies on hand strength have concentrated on handgrip. Human beings are often required to use the fingers for fine control. The human hand is important whether performing casual activities or professional/industrial tasks. The surgeon and construction worker are examples in the latter categories. In the science of ergonomics, not many studies have focused on the performance of human fingers. However, Imrhan and Loo (1989) discussed finger strengths in various age groups, while Imrhan (1987) produced one of the few studies on hand torque strength.

Individuals use their fingers to turn things every day. They turn the keys in the ignition of their cars. They also lightly tighten and loosen screws when they have to change a flat tire or when they turn the knob of a lamp to read interesting literature, and much more. The need to turn things with one's fingers is important where the object is too small to be turned with the hand and where the use of tools is inappropriate, either because of limited space for their manipulation or because the object cannot be properly grasped by tools. In some factories, there exists sensitive equipment that cannot be manipulated by any other tool but the fingers. Such equipment is seldom used throughout factories due to the repetitiveness of the activity, which may cause cumulative trauma disorders. Examples of this would be when the worker constantly handles small screws, placing them in a certain position, lightly loosening and tightening them for reworking the product.

In order to understand fully the role of fingers in the hand, researchers need data not only on finger pull strengths but also on finger torque strengths for particular segments of the population. The purpose of this study was to investigate the nature of such force strength.

2. METHODS

The main objective of the study was to determine the magnitude of finger torques while turning handles of different sizes. Handles consisted of seven different diameters (2.0 cm, 1.25 cm, 1.0 cm, 0.75 cm, 0.5 cm, 0.25 cm and 0.15 cm) and of three surfaces (smooth, knurled, or flanged). The study was conducted with 23 subjects of an assembly line in a local industrial plant. Fifteen worked in a different shift from the others. Permission for testing subjects was granted by the University of Texas at Arlington's (UTA) Human Re-

search Review Committee, which reviewed the study's proposal for risks to subjects and other concerns, and by the management of the plant.

This investigation was divided into two stages. The first stage was to perform all measurements clockwise, while in the second stage all measurements were counterclockwise. Each test group consisted of 2 to 3 individuals. Subjects performed 2 to 3 tests per session, with a rest period of 2 to 3 minutes between tests. Such constraints were imposed to avoid fatigue and negative effects on the results.

2.1. Stage 1

2.1.1 Subjects

The participants in this investigation can be described as a voluntary sample of 10 males and 13 females in an assembly line, between the ages of 25 and 55 years. Table 1 describes the surface means and standard deviations of all 23 subjects. The subjects were from various nationalities. The majority of the subjects were right-handed.

SURFACE	N	MEAN	STANDARD DEVIATION	MINIMUM	MAXIMUM
Smooth	322	29.24	31.43	0	169
Knurled	230	61.57	39.37	8	166
Flanged	138	58.27	27.16	18	139

Table 1. Summary of Means and Standard Deviation of All 23 Subjects

2.1.2 Apparatus

A torque measuring device (Dillon Model T. Static) was connected to a digital torque indicator (Dillon Model AFTI) to measure and display torques generated by subjects, in units of inch ounce. The handles were made from aluminum and were attached to an adapter on the shaft of the torque tester. This allowed the different sizes of handles to be exchanged easily (Figure 1).

Figure 1. Torque Measuring Device and Digital Torque Indicator

2.1.3 Strength tests

Each handle was of a different size and surface type. For the knurled and smooth handles, diameters ranged from 0.15 cm - 2.0 cm. For the flanged, the widths were 1.5 cm, 1.25 cm and 1.0 cm (Figure 2).

These surfaces corresponded to types and sizes of different tools (handles) found in an assembly line. In all cases, subjects held these tools in the same way during the tests (Figure 3). These tools were held firmly and stabilized with one hand. The tools were then turned either clockwise or counterclockwise, depending on instructions from the tester, without slipping or jerking. Subjects built up to the maximum gradually in about 1 to 2 seconds and held it for about 3 seconds.

Figure 2. Diameters of Handles Used for This Experiment

2.1.4 Torque direction

Two torque directions were tested – clockwise and counterclockwise. For clockwise torque, the tester switched direction on the torque indicator to measure this directional torque. Subjects held their left hand (if they were right handed) or vice versa, on the measuring apparatus (Figure 3) for more stability. Measurements on each size and type were taken twice (variable combination). If the two measurements were not within 15% of each other, the measurement was repeated. The larger value was used for data analysis.
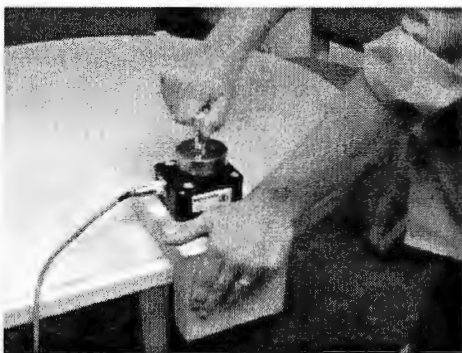
Figure 3. Finger Position on Device Measurement

2.1.5 Procedure

Subjects were tested in groups of 2 to 3. Resting time between trials for any subject was enough after 2 to 3 minutes to minimize fatigue and injuries. Subjects were tested as workload allowed. Subjects were tested at an empty desk near the production line. The testing method was designed to avoid disruption of work on the assembly line.

3. RESULTS AND DISCUSSION

Data obtained through this investigation was analyzed using the Statistical Analysis

System (SAS). Analysis of Variance (ANOVA) and Tukey's Test were the main analytic methods to draw conclusions about this study. The ANOVA was performed on the strength data (Torque) with surface, direction and diameter as independent variables.

Table 2 shows the means for the various combinations of diameter, surface type and direction of torqueing. This data reflects the difference in the task and handle designs. The overall mean when tightening any surface is 48.63 inch ounce with a standard deviation of 37.03 inch ounce, while for loosening it is 43.01 inch ounce with a standard deviation of 36.69 inch ounce. This indicates that subjects have greater finger torque strength when tightening by a considerable amount than when loosening small handles.

Surface / Type (diameter in cm)		Smooth		Knurled		Flanged	
		Tighten	Loosen	Tighten	Loosen	Tighten	Loosen
2.00	mean	89.74	86.87	116.39	119.95	88.00	73.74
	Std. Dev	28.73	33.69	22.27	31.44	25.15	25.69
1.25	mean	52.52	43.26	87.04	74.26	59.04	51.00
	Std. Dev	17.10	17.02	19.11	22.66	21.02	19.38
1.00	mean	31.39	26.17	62.91	48.69	39.61	38.21
	Std. Dev	10.66	9.62	13.37	17.07	16.85	13.57
0.75	mean	17.26	16.74	36.65	23.78	n/a	n/a
	Std. Dev	6.47	8.64	11.70	6.88	n/a	n/a
0.50	mean	12.57	12.13	26.48	19.52	n/a	n/a
	Std. Dev	5.15	5.35	7.83	6.84	n/a	n/a
0.25	mean	7.09	7.13	n/a	n/a	n/a	n/a
	Std. Dev	4.38	5.11	n/a	n/a	n/a	n/a
0.15	mean	2.87	3.65	n/a	n/a	n/a	n/a
	Std. Dev	2.55	3.38	n/a	n/a	n/a	n/a

Table 2. Summary of 7 Handles and Different Types of Surfaces in Inch Ounce

Analysis of Variance was performed on the data to determine which of the torqueing factors (diameter, surface, and direction) had any effect on torque strength. The results indicate that the torque was influenced significantly (P-value smaller than 0.05) by all three factors and by the interaction of surface and diameter [F(surface) = 285.73; p=0.0001]. This interaction demonstrates that torqueing trends on knurled handles and smooth handles were different.

On the other hand, the results on the interaction between surface and direction as well as diameter and direction do not influence torqueing significantly (P-value larger than 0.05). In other words, the interaction of direction with surface [F(2.56); p=0.0784] and with diameter [F(0.56); p=0.7648] does not have major implications on torqueing strength. The ANOVA table for all surfaces, directions and diameters appears in Table 3.

In order to compare the difference between surfaces, Tukey's test was used. This procedure allows paired comparisons that "sustain the credibility of the type I error rate)" (Walpole, et. al., 1998). Tukey's test (Table 4) confirms that torqueing strength difference is small between the knurled and flanged surfaces (difference between means = 3.301 inch ounce). For P-value smaller than 0.05 these surfaces are not significant. Additionally, the mean torqueing strength between the knurled and smooth surfaces is larger (difference between means = 32.327 inch ounce) than between the flanged and smooth surfaces (difference between means = 29.026 inch ounce). This means that knurled surface strength was 89.8% greater than smooth surface strength, while flanged surface strength was 88.6% greater than smooth surface strength.

ANOVA TABLE

Dependent Variable: Torque

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	23	745471.21	32411.79	110.96	0.0001
Error	666	194545.22	292.11		
Corrected Total	689	940016.43			

	R-Square		C.V.	Root MSE	TORQUE Mean
	0.79		37.298	17.09	45.82

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Surface	2	166927.86	83463.93	285.73	0.0001
Type	6	555464.64	92577.44	316.93	0.0001
Direction	1	5465.74	5465.74	18.71	0.0001
Surface x Type	6	15144.19	2524.03	8.64	0.0001
Surface x Direction	2	1492.94	746.47	2.56	0.0784
Type x Surface	6	975.83	162.64	0.56	0.7648

Note: Signif Interaction
No Signif Interact

Table 3. ANOVA Table for Finger Torque Strength in Inch Ounce

TUKEY'S TEST

Alpha = 0.05   Confidence = 0.95   DF = 666   MSE = 292.1099
Critical Value of Studentized Range = 3.322

SURFACE COMPARISON	SIMULTANEOUS LCL	DIFFERENCE BETWEEN MEANS	SIMULTANEOUS UCL	Note
Knurled - Flanged	-1.021	3.301	7.624	
Knurled - Smooth	28.861	32.327	35.793	***
Flanged - Knurled	-7.624	-3.301	1.021	
Flanged - Smooth	24.941	29.026	33.111	***
Smooth - Knurled	-35.793	-32.327	-28.861	***
Smooth - Flanged	-33.111	-29.026	-24.941	***

Note: *** Comparisons significant at the 0.05 level

Table 4. Tukey's Test Surfaces Paired Comparisons in Inch Ounce

These results also support the idea that greater handle contact area provides "greater tangential force, and hence, greater tangential torque" (Imrhan and Farahmand, 1991). For example, the larger the diameter in a smooth surface the greater the torqueing strength (80.87 inch ounce), while the smaller the diameter with the same surface, the smaller the torqueing strength (3.26 inch ounce) (Figure 4).

For a large handle diameter, the fingers are most likely to achieve optimal leverage. "Assuming a constant coefficient of limiting friction for these exertions" (Imrhan and Farahmand, 1991), the torqueing strength decreases as the handle diameter decreases.

Figure 5 expresses graphically the interactions between surfaces and torque strengths with respect to direction. When tightening or loosening handles with the smooth surfaces, the difference in torqueing is not as great as when using knurled or flanged surfaces. For smooth surfaces tightening a handle was measured at 30.49 inch ounce, while loosening a

handle measured 27.99 inch ounce, a small difference of only 2.5 inch ounce. On the other hand, knurled surfaces revealed a difference between the two directions tested of 8.65 inch ounce, and flanged surfaces had a difference of 7.90 inch ounce.
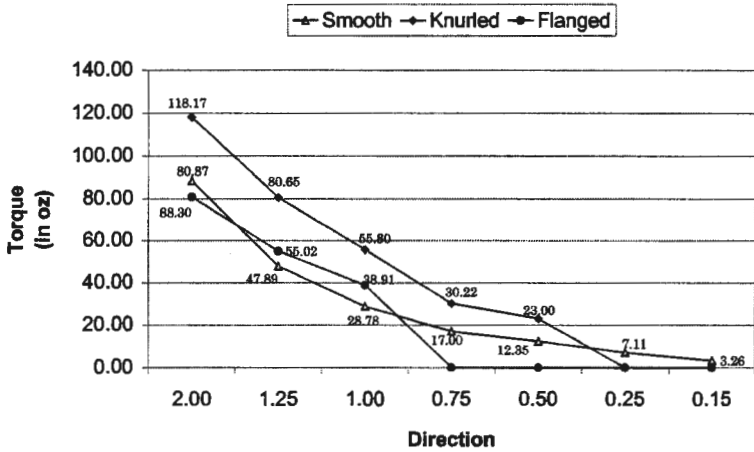
Comparison between Surface and Direction

Figure 4. Different Diameters of Handles Compared to Three Different Surfaces

Data Interactions

Figure 5. Torqueing Strength with Different Surfaces
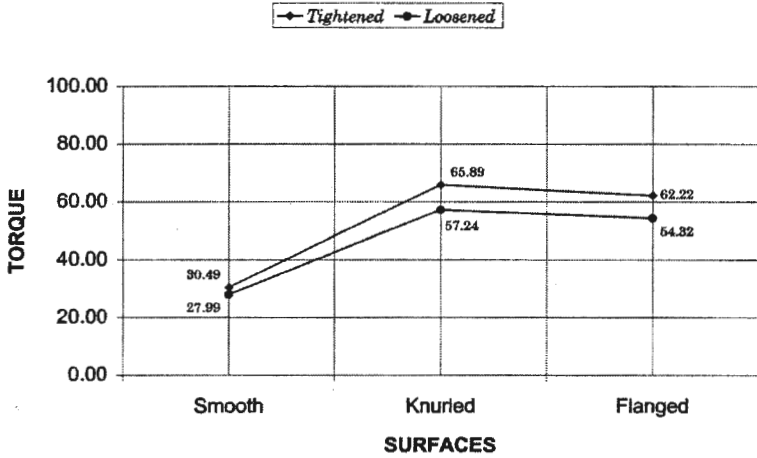
Table 5 shows the means for surface and diameter and confirms the strong interaction between these two variables. This study shows that the bigger the diameter, the stronger the finger torque will be, and as the diameter decreases so will finger torque. However, the strength–diameter difference is not the same for smooth and knurled surfaces (significant interaction).

Surface Type (diameter in cm)		Smooth	Knurled	Flanged
2.00	mean	88.30	118.17	80.87
	Std. Dev	31.00	27.00	26.15
1.25	mean	47.89	80.65	55.02
	Std. Dev	17.51	21.71	20.40
1.00	mean	28.78	55.80	38.91
	Std. Dev	10.38	16.78	15.14
0.75	mean	17.00	30.22	n/a
	Std. Dev	7.55	11.50	n/a
0.50	mean	12.35	23.00	n/a
	Std. Dev	5.19	8.07	n/a
0.25	mean	7.11	n/a	n/a
	Std. Dev	4.71	n/a	n/a
0.15	mean	3.26	n/a	n/a
	Std. Dev	2.98	n/a	n/a

Table 5. Comparison of Surfaces versus Handle Diameters in Inch Ounce

4. CONCLUSIONS

The torqueing factors (surface, diameter and direction) had an effect on torque strength, as did the interaction between surface and diameter (P-value = 0.001). Additionally, as Figure 5 shows, tightening and loosening motions on a smooth surface had a smaller difference than on the knurled and flanged surfaces. Subjects are stronger when tightening knurled and flanged surfaces than when loosening them. Hence, handles for torqueing should be designed to take advantage of these effects. Smooth surfaces at small diameters were as good as rough surfaces for torqueing, but at large diameters knurled surfaces were better.

In conclusion, measuring torqueing under three different conditions and factoring for each condition reveals different torqueing strengths. Therefore, in designing or assigning tools to workers, we must take into consideration the effects of these factors. For example, a handle with a larger diameter requires less musculoskeletal effort than a smaller diameter handle. Further investigation is necessary to confirm if there exists any difference in finger torqueing when using different finger positions.

5. ACKNOWLEDGEMENTS

6. REFERENCES

Armstrong, T. J., Foulke, J. A., Joseph, B. S. and Goldstein, S. A., 1982. Investigation of cumulative trauma disorders in a poultry processing plant. *American Industrial Hygiene Association Journal*, **43**, 103-116.

Imrhan, S. N. and Farahmand, K., 1991. Handle design parameters and manual torquing in simulated oil and gas extraction tasks. *Advances in Industrial Ergonomics and Saftey III*, 579-585.

Imrhan, S. N. and Jenkins, G. D., 1990. Hand turning torques in a simulated maintenance task. *Advances in Industrial Ergonomics and Safety II*, 437-444.

Imrhan, S. N. and Loo, C., 1986. Torque capabilities of the elderly in opening screw top containers. *Proceedings of the Human Factors Society*, 1167-1171.

Imrhan, S. N. and Loo, C., 1987. Some practical considerations in selecting lids for testing wrist-twisting strength. *Proceedings of the IXth ICPR*, 706-711.

Imrhan, S. N. and Loo, C., 1989. Trends in finger pinch strength in children, adults, and the elderly. *Human Factors*, **31**(6), 689-701.

Imrhan, S. N. and Sundararajan K., 1992. An investigation of finger pull strengths. *Ergonomics*, 1992, **35**, 289-299.

Nagashima, K. and Konz, S., 1986. Jar lids; effect of diameter, gripping materials and knurling. *Proceedings of the Human Factors Society*, 30th Annual Meeting, 672-674.

Walpole, R. E., Myers, R. H., Myers, S. L., 1998. *Probability and Statistics for Engineers and Scientists*, 377-380, 394-396, 481-483.

Jewel Compression: Advanced Data Compression Algorithm

José R. Espino

Mentor: Dr. Arthur A. Reyes

Department of Computer Science and Engineering

ABSTRACT

Traditional methods of data compression employ fixed-length, block-by-block compression. These methods assume the input data stream is of known length with a defined beginning and end. Unfortunately, these methods of data compression are not effective for compressing continuous, real-time data streams. The Jewel Compression Algorithm is a lossless, adaptive, time-dependent compression technique for compressing variable length data streams. Using a functional programming approach, a Jewel Compression prototype was developed that simulated the compression of continuous, real-time data streams.

1. INTRODUCTION

Data compression refers to the process of symbolically representing a given unit of information in a more compact form, in essence, one that requires less storage space. This technique of reducing storage space is one of the most important and researched ideas in computer science due to its practical applications. As the use of computers proliferates, so does the data that is transmitted over intranets, satellites, audio communication channels, and, of course, the Internet. If this data, whether it be medical images, data records, military information, or data used in entertainment, were to be transmitted in its original size, bandwidth would suffer. Today much of the data transmitted over intranets and the Internet is compressed. Executable files are compressed using popular compression programs such as WinZip™. Most graphics are transmitted in a format that reduces storage space, and audio and video are usually compressed. Thus, modern society's casual dependence on data has caused the process of data compression to become ubiquitous and transparent in every day life.

Data compression relies on four traditional methods: run length encoding, statistical methods, dictionary-based methods, and lossy image compression [CRW87]. These four approaches to data compression share one commonality—they operate on fixed-length files (files that have a specified beginning and end). Since much of data exists in the form of fixed-length files, these compression techniques have been developed and refined. Unfortunately, these methods of data compression are incapable of operating on continuous streams of data.

The U.S. Navy maintains a diverse fleet of airborne vehicles that utilize digital avionic systems. Due to monetary constraints and the diversity of the software systems, it is impractical to develop brand-new software for these avionic systems every time a change must be incorporated. Also, much legacy code must be adapted to work with new aircraft and equipment that is incorporated into the existing fleet. Regression testing then becomes mandatory to determine whether the new system is operating according to specified requirements.

The Naval Air Warfare Center/Aircraft Division (NAWC/AD) proposed Jewel Compression (JC) as a means for compressing continuous data streams and, ultimately, for automating regression testing. The JC algorithm is a lossless, time-dependent compression technique for compressing data streams that are of a real-time nature [WY99]. Given a software system that is to undergo regression testing, JC will accept the test unit's continuous output

data stream, partition the data into manageable units, sort the units, compress the data, and finally restore the compressed data to its original temporal order.

This report is divided into five sections. The first section is an introduction. Section 2 describes the basic operational scheme of the JC system. A discussion of the implementation and design decisions is presented in Section 3. Experimental results are provided in Section 4. Finally, Section 5 presents a list of conclusions as well as future development of the JC system.

2. BASIC SCHEME

The JC system consists of seven components including an input stream, a demand process that consumes the compressed output of the JC system, a Pack module that partitions the input stream into units, a Sort module for lexicographical sorting, a Compress module which performs symbolic compression, and an Unpack module used to separate the data from the control character sequences [WY99]. Control character sequences are sequences of characters used to denote whether the producer process has terminated. The seventh component is an executive module named the Jewel Compression Controller (JCC) that decides when data is added to the Jewel Data Structure (JDS), compression opportunities, and other modifications of the JDS. A JDS contains the data that may be compressed or uncompressed as well as any existing compression definitions. A simple example will be presented by describing an input stream of lower case, printable characters (ASCII symbols 61H – 80H) and how the JC processes it.

For a simplified example, let us assume that we have a data stream that is finite. This stream consists of a sequence containing the five distinct characters 'a', 'b', 'c', 'd', and 'e', which may be repeated any number of times. Now, let the stream of data be represented as a list of the following form.

$$[a,b,c,d,a,c,e,a,b,d,c,e,f]$$

The Pack module has one purpose: it accepts a continuous data stream and partitions the incoming input stream into manageable units referred to as clumps. It does this by building a dictionary of hash keys, which are used to signify the beginning of a clump of data. Every time the Pack module encounters a character that has not been seen before, it decides whether the newly seen character should become a hash key. The decision is based on a user specified probability. Pack automatically converts the first character of the input data stream into a key so that it can begin partitioning the data without loss. Partitioning is carried out by examining each input character. If the character is not a key, it is added to the current clump; otherwise, the current working clump is finished, and a new clump is created. In this example, I will use two hash keys K1 and K2, where K1 = a and K2 = d, which are in the set of characters that comprise the input stream. I have assumed that Pack randomly chose 'a' and 'd' to be keys. Pack then operates on the stream of data and partitions it into following list of lists. Notice that each clumps begins with one of the hash keys ('a' or 'd').

$$[[a,b,c],[d],[a,c,e],[a,b],[d,c,e,f]]$$

Once the data is partitioned, the Jewel Compression Controller (JCC) is invoked and the partitioned data is made available to it as input. The JCC is responsible for controlling the sort, compress, and unsort activities, creating and maintaining the JDS, as well as man-

aging the demands of the consumer process. In order to simplify the example, I assume the consumer process presents the JCC with enough wait demands that allow JCC to sort, compress, and unsort the input data stream. In Section 3 I will discuss under what conditions JCC fills the JDS with clumps.

Sort reorders the list of clumps lexicographically in preparation for compression, but this destroys the temporal order of the input data stream. It is the responsibility of the JCC to ensure that when a clump is added to the JDS the original temporal order to the data stream can be reconstructed. JCC ensures that this restriction is met by adding data to the JDS in the form of pairs where the first element represents the current clump and the second element is a marker that is used for temporal restoration. Markers are always the second element of the 2-tuple. I refer to this position in the 2-tuple as the Right-Hand-Side (RHS). The first element of the 2-tuple always contains data and is referred to as the Left-Hand-Side (LHS). For the initial model, the RHS component is a copy of the next clump in the Pack input data stream. When adding a clump I define the following policy: A clump may be added to the JDS as long as its RHS element does not appear as a LHS element in any data pair in JDS. A policy of this nature guarantees that the original, temporal order can be unambiguously recovered. This will be referred to as a no-harm-add policy. In my example, the input list would be processed into a JDS of the following form where [eof] represents an end-of-file marker (note that the clump symbol eof is assumed to not appear in the input data stream).

[([a,b,c],[d]), ([d],[a,c,e]), ([a,c,e],[a,b]), ([a,b], [d,c,e,f]), ([d,c,e,f], [eof])]

After Compress module operates on the JDS, it returns the following list sorted lexicographically from least to greatest based on the first element in each pair. In my example, after a call to Sort, the list would be as follows.

[([a,b], [d,c,e,f]), ([a,b,c],[d]), ([a,c,e],[a,b]), ([d],[a,c,e]), ([d,c,e,f], [eof])]

Note that by probing the RHS element of each 2-tuple it is possible to reconstruct the temporal order of the input sequence. It is necessary to find the LHS of a pair that does not match the RHS of any pair in the input data stream. The non-matching element signals the last clump and the input data stream can be restored.

Next, the Compress module creates an Entropy Matrix (EM) to compute compression opportunities. The EM is used to record matches between different data clumps and is a degenerate matrix. Figure 1 shows the EM for this example. The far left column denotes the length of match, that is, how many clumps match to a length of two, three, etc. The other columns denote how many clumps match for a given length. Columns represent disjoint sets of clumps. The EM for my example is detailed below and is read as two entries (clumps) match to length two.

2	2	0	0	0
3	0	0	0	
4	0	0		
5	0			

Figure 1. Entropy Matrix 1

In this example, this represents a saving of at least two bytes.

Compression will now take place as the Compress module replaces each joint occurrence of the characters 'a' and 'b' with the new symbol **X**. After compression, the JDS now contains the following data.

[([X], [d,c,e,f]), ([X,c],[d]), ([a,c,e],[X]), ([d],[a,c,e]), ([d,c,e,f], [eof])]

The data stream is out of temporal order and this is where the unsort function plays a key role. It restores the clump stream to its original temporal order and JCC delivers the oldest compressed clump upon demand. In my example, the compressed data restored to its original, temporal order is detailed below.

[[X,c],[d],[a,c,e],[X],[d,c,e,f]]

3.   IMPLEMENTATION

The implementation is divided into four sections. Three of these sections explain the implementation and design decisions for the Pack, Sort, and Compress modules, and the other section provides a description of the JDS and the JCC. The JC system prototype has been developed as a sequential (i.e., "clocked") model.

3.1.   Pack

Raw data is input into the JC system and it is the job of the Pack module to partition the data into manageable units. As data enters the JC system, it is in a non-partitioned state: it is simply a sequence of bytes that were stored in the file. The Pack module must partition the data and it does so by using hash keys, which are picked from the input stream. The Pack module maintains a table of byte-values that have been encountered in the data stream. Each time a new byte value is encountered, the Pack module decides whether the newly encountered byte-value should become a hash key. The decision is based on a user-specified probability. Byte-values have a one-time chance of becoming hash keys. If they are not made into hash keys on their first occurrence, they are no longer eligible for consideration on their next appearance. Clumps always begin with a hash key and may only contain one hash key character per clump.

With all partitioning probabilities less than 100% there is a chance that the first byte-value in the data stream will not be converted to a hash key. This situation would result in the Pack module discarding the data at the beginning of the input stream until a hash key is created. Furthermore, the data loss problem would be exacerbated by the probability value—a lower probability value would result in more data loss. To prevent data loss, the first value encountered in the data stream is always converted into a hash key regardless of the partition probability value specified by the user. This condition will hold even if the user specified a partitioning probability of zero. Also, converting the first byte-value into a hash key is a necessary condition to begin partitioning the data since the partition process must begin on the first byte-value.

I have provided the source code for the body of the pack function: build_list whose responsibility is to build the list of clumps, and build_clump, which creates the clumps. The source for the add function is not included. This function simply appends the first parameter to the second parameter.

All source code provided in this report is Erlang code. The arrow symbol (->) denotes a return value from a function. Lists are enclosed in braces (i.e., [List]), the empty list is represented by [], and the prepend operator in lists is the | symbol. All tuples are enclosed with curly braces (i.e., {e1,e2}).

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                              pack.erl
%
%       A file specified by Name is bound to Input. Keys is dictionary
%       of hash keys. Return contains a list of clumps and the total
%       count of clumps.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

pack( Prob, Name ) ->
      Input = process( Name ),
      Keys = build_dict( Input, Prob ),
      Return = build_list( Input, [], 0, Keys ).


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%       Function build_list creates a list of clumps. Return is a
%       2-tuple in the form of { Unprocessed_Data, Clump_List }. Note
%       that Return is created bound by a call to build_clump which
%       operates on the unprocessed list eventually consuming it
%       and thus providing for function termination. At each recursive
%       invocation of build_list, the remainder of the unprocessed list
%       is used and the clump count is increased.
%
%       Case 1:     The non-partitioned list is empty. Building the list
%                   of clumps is finished. Return the list of clumps and
%                   the total of clumps in the list.
%       Case 2:     The non-partitioned list contains one clump. Process
%                   the last clump.
%       Case 3:     The non-partitioned list contains multiple elements.
%                   Use build_clump to process as much of the non-
%                   partitioned data then invoke build_list on remaining
%                   non-partitioned data.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
build_list( [], Output, Count, Keys ) ->
      { Output, Count };

build_list( [H], Output, Count, Keys ) ->
      Return = build_clump( [H], Output, [], false, Keys ),
      build_list( element( 1, Return ), element( 2, Return ), Count + 1,
                  Keys );

build_list( [H|T], Output, Count, Keys ) ->
      Return = build_clump( [H|T], Output, [], false, Keys ),
      build_list( element(1, Return ), element( 2, Return ), Count + 1,
                  Keys
                  ).
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%       IsKey is bound to a boolean value by searching the Keys dictionary
%       for the current head value of the non-partitioned list. It is
%       necessary to test for three conditions.
%
%       1:      The head element is a key, but it has not been seen while
%               building the current clump. Start a new clump with the key
%               at its head.
%       2:      The head element is not a key. Add it to the current
%               working clump.
%       3:      The head element is a key and has been seen in the working
%               clump. This signals the beginning of a new clump. Do not
%               add the head element to the working clump. Instead, the
%               working clump is added to the list of clumps.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
build_clump( [], Output, Temp, Seen, Keys ) ->
     { [], add( Temp, Output ) };

build_clump( [H], Output, Temp, Seen, Keys ) ->
     IsKey = hd( dict:fetch( H, Keys ) ),
     if
          IsKey == true, Seen == false -> build_clump(    [],
                                                          Output,
                                                          add(H,Temp),
                                                          true,
                                                          Keys
                                                     );
          IsKey == false -> build_clump(         [],
                                                 Output,
                                                 add(H,Temp),
                                                 true,
                                                 Keys
                                            );
          IsKey == true, Seen == true -> { [H], add( Temp, Output ) }
     end;

build_clump( [H|T], Output, Temp, Seen, Keys ) ->
     IsKey = hd( dict:fetch( H, Keys ) ),
     if
          IsKey == true, Seen == false -> build_clump(    T,
                                                          Output,
                                                          add(H,Temp),
                                                          true,
                                                          Keys
                                                     );
          IsKey == false -> build_clump(         T,
                                                 Output,
                                                 add(H,Temp),
                                                 true,
                                                 Keys
                                            );
          IsKey == true, Seen == true -> {[H|T], add( Temp, Output )}
     end.
```

3.2. Jewel Data Structure/Jewel Compression Controller

The JDS is a pair where the LHS element of the JDS is a list of clump pairs. The clump pairs are structured in the following form: (Data, Marker). Data elements represent data that will participate in compression activity and the Marker is used to reconstruct the temporal order of the input data stream. The JDS RHS is a list of pairs where the pairs are compression definitions. The first element is the compression definition and the second element indicates what data should be used to replace the compression symbol whenever the compression must be undone. I have provided a description of a JDS where the symbol [H|T] represents a prepend operation (i.e. prepends an element H onto a list of elements T).

JDS = [([(Data, Marker) | Other_Clumps], [(Defs, Chars) | Other_Defs])]

Now that I have defined the JDS it is necessary to describe the JCC whose responsibilities include maintenance of the JDS, sorting, compressing, restoring the temporal order of the input data, and managing the consumer process's demands. A data flow diagram for JCC is depicted in Figure 2.
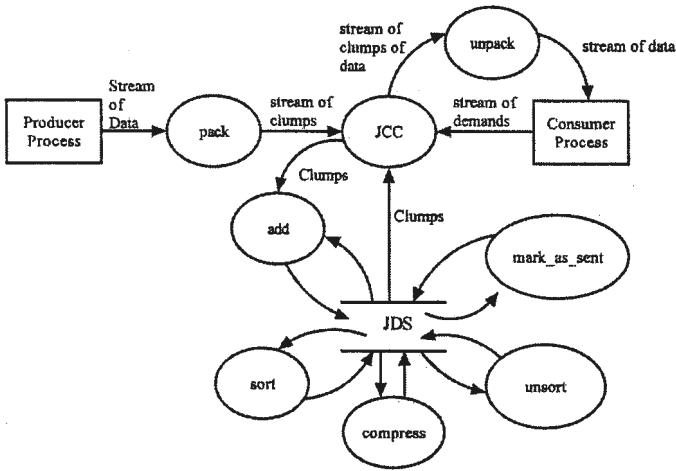
Figure 2. JC Data Flow Diagram

Discussion of JCC is broken down into several categories. First I will provide an overview of JCC, its interface, and how a synchronicity is simulated. A discussion of the primitive **mark_as_sent** follows the overview. Finally, I will present the 18 different cases contained in JCC and group them by category. There are a total of seven categories: Base cases which result in termination of JCC execution, operations on an empty input list and empty JDS, operations on an empty JDS, exceptions, unconditional compression, cases where JCC must use the primitive **mark_as_sent**, and operations on non-empty input lists and a non-empty JDS.

Eighteen possible scenarios exist for a given execution of the JC system, thus JCC can be abstractly viewed as a module that processes eighteen unique cases. A quick examination of the three parameters required by JCC's interface, an input clump list, a JDS, and a list of the consumer process demands, will help to explain why eighteen unique cases exist. The first parameter, the input list, can be empty, can contain one element, or can contain multiple elements for a total of three states. A JDS can either be empty or non-empty for a total of two states. The consumer process may present JCC with an empty demand list, a list with a wait demand followed by other demands, or a send demand followed by other demands (three more states). When the three parameters and their independent states are considered, simple math says that there are eighteen possible combinations.

A consumer process presents JCC with a list of demands composed of either wait demands or send demands. Each demand can be viewed as one cycle of execution during which JCC can terminate execution, add data to the output stream, or engage in compression.

Wait demands are simply cycles during which the consumer process does not request a data item. Wait cycles give the JCC opportunities to modify the JDS. Send commands cause JCC to transmit data to the consumer process. When the consumer process presents a send demand, JCC uses the primitive **mark_as_sent** to remove the oldest data pair from the JDS. The term oldest refers to the first element in the LHS of the JDS. Once the data pair is removed from the JDS it is added to the output stream and is no longer eligible to participate

in further compression activity. This is an assumption that holds only for this initial model. Later models will relax this assumption. If the sent data is compressed, a compression definition is sent to the consumer process as well. The compression definitions will precede compressed clumps.

The JCC has six base cases that cause termination of execution (cases 1, 4, 7, 10, 13, 16). All of them share one common characteristic: JCC is invoked with an empty demand list. The empty demand list represents a consumer process that has terminated. Once the consumer process terminates, it is not necessary for JCC to process anymore input or continue execution. The six base cases occur when JCC is invoked with the following parameters summarized in Table 1.

Parameter 1: Input List	Parameter 2: JDS	Parameter 3: Demand List	Action
Empty	Empty	Empty	Terminate
Single Element	Empty	Empty	Terminate
Multiple Elements	Empty	Empty	Terminate
Empty	Non-Empty	Empty	Terminate
Single Element	Non-Empty	Empty	Terminate
Multiple Element	Non-Empty	Empty	Terminate

Table 1. Summary of JCC Base Cases

The Erlang code for each of the six base cases is provided below. Code will be provided for all the cases on a category-by-category basis throughout the rest of this section.

```
jcc( [], [], [] ) ->
        [];

jcc( [C], [], [] ) ->
        [];

jcc( [C1,C2|Clumps], [], [] ) ->
        [];

jcc( [], [{O,N}|T], [] ) ->
        [];

jcc( [T], JDS, [] ) ->
        [];

jcc( [C1,C2|T], JDS, [] ) ->
        [];
```

The second category of cases covers JCC operations where the input list and JDS are empty, but the consumer process presents the JCC with demands (cases 2, 3). A consumer process can present the JCC with a send demand or a wait demand. In each case, JCC will output a special marker that signifies that the producer process (the input list) has been completely consumed. JCC will continue to process the demand list until the demand list is empty. Each consumed demand will result in a producer-terminated marker output.

```
jcc( [], [], [W|Dmds] ) ->
        lists:append( [['P', 'T']], jcc( [],[], Dmds ) );

jcc( [], [], [S|Dmds] ) ->
        lists:append( [['P', 'T']], jcc( [], [], Dmds ) );
```

Categorization by operations on an empty JDS yields three different cases (cases 5, 6, 9). The first case occurs when the input list contains one clump and the consumer processes demand that JCC wait. This wait demand gives JCC time to add the last clump of data to JDS by constructing a 2-tuple (Last Data, EOF), where EOF is an end-of-file marker (that does not appear in the input alphabet). On the other hand, the consumer process could present JCC with a send demand. Then JCC would append the final clump to the output stream and invoke itself with an empty input list. The third is like the second case with one difference: the input list contains multiple clumps. JCC appends the first clump to the output stream and invokes itself with the remaining input list and demand list.

```
jcc( [C], [], [W|Dmds] ) ->
        jcc( [], [{C, eof}], Dmds );

jcc( [C], [], [S|Dmds] ) ->
        lists:append( [C], jcc( [], [], Dmds ) );

jcc( [C1,C2|Clumps], [], [S|Dmds] ) ->
        lists:append([C1], jcc( lists:append( [C2], Clumps ), [], Dmds ));
```

JCC requires that JDS be in temporal order at each invocation of JCC. This requirement coupled with the no-harm-add policy creates an exception case that will cause premature termination of the program (case 8). This situation can occur when JCC is invoked on an input list with multiple clumps that are equal, an empty JDS and a wait demand. In order to understand this behavior, it is necessary to examine the algorithm used to restore the temporal order of the data.

```
jcc( [C1,C2|Clumps], [], [W|Dmds] ) ->
        if
                C1 < C2 -> jcc( [C2|Clumps], [{C1,C2}], Dmds );
                C1 > C2 -> jcc( [C2|Clumps], [{C1,C2}], Dmds );
                C1 == C2 -> { error, consecutive_clump_exception }
        end;
```

Unsort is the function used to restore the temporal order. From an abstract point of view, unsort considers the data pairs as nodes in a directed, acyclic graph. It locates the node that has no exiting directed arc and considers it to be the newest data pair. Unsort does this by locating a right-hand element that does not appear anywhere on the left-hand side. This is considered to be the youngest node. It then reconstructs the temporal order with the head of the data list being the oldest data and the last element the youngest. In Figure 3 below, I have provided a directed, acyclic graph diagram representing the data contained in the JDS based on the example presented in Section 2.
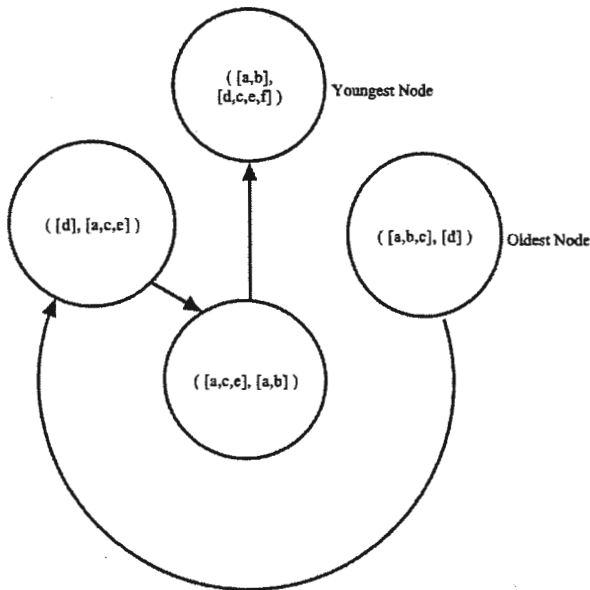
Figure 3 Graph Representation of Clump Pairs

If a repeated element were to be added to the graph, a cycle occurs and the algorithm will be unable to reconstruct the temporal order. When this situation occurs, JCC terminates prematurely. This problem is due to a bad design decision and a solution will be discussed in Section 5.

An opportunity for unconditional compression activity also exists (case 11). By unconditional I mean that this is the only activity that is available during the wait cycle. This occurs when JCC is invoked on an empty input list, a non-empty JDS, and the consumer process presents JCC with a wait demand. During this cycle, all input has been consumed and JCC is forced to engage in compression activity to avoid wasting a clock cycle.

```
jcc( [], JDS, [W|Dmds] ) ->
        JDSS = sort( JDS ),
        JDSC = compress( JDS ),
        JDSU = unsort( JDS ),
        jcc( [], JDSU, Dmds );
```

The next category of cases is based on the use of the primitive **mark_as_sent** (cases 12, 15, 18). Three cases are contained in this category and all occur when the consumer process presents JCC with a send demand. All three cases involve JCC operations on a non-empty JDS as well. When the input clump list is considered, it can be in three states: empty, non-empty with one element in it, or non-empty with multiple elements in it. In all three cases JCC does not modify the input list. It uses **mark_as_sent** to remove the oldest element from JDS, appends it to the output clump stream, removes the send demand from the consumer demand list, and continues execution.

```
jcc( [], [ {[{0,N}|T], Defs} ], [S|Dmds] ) ->
      lists:append(    [0],
                       jcc(   [],
                              mark_as_sent(0,[{{0,N}|T},Defs}]),
                              Dmds
                          )
                   );

jcc( [T], [ {[{0,N}|T], Defs} ], [S|Dmds] ) ->
      lists:append(    [0],
                       jcc(   [T],
                              mark_as_sent(0,[{[{0,N}|T],Defs}]),
                              Dmds
                          )
                   );

jcc( [C1,C2|T], [ {[{0,N}|T],  Defs} ], [S|Dmds] ) ->
      lists:append(    [0],
                       jcc(   [C1,C2|T],
                              mark_as_sent(0, [ {[{0,N}|T], Defs} ] ),
                              Dmds
                          )
                   );
```

Finally, the seventh category is comprised by operations on a non-empty input list and a non-empty JDS (cases 14, 17). Both cases occur when the consumer process presents JCC with a wait demand. When the consumer presents the wait demand, if the input list contains one element, a pair in the form (Data, end-of-file) is created and unconditionally added to JDS. This is a safe decision since the end-of-file marker will never appear in the left-hand side entries and, thus, does not violate the no-harm-add policy. The second case occurs when the input list contains multiple clumps. First, JCC determines whether it is safe to add a clump to JDS. If it is safe, the clump is added to JDS, and JDS is compressed; otherwise, the clump is not added and JDS is not compressed.

```
jcc( [T], JDS, [W|Dmds] ) ->
      jcc( [], lists:append( [JDS], [{T,eof}] ), Dmds );

jcc( [C1,C2|Clumps], JDS, [wait|Dmds] ) ->
      Flag = can_add( C1,C2, JDS ),
      if
          Flag == true -> jcc( [C2|Clumps],
                               unsort(
                                  compress(
                                    sort(lists:append([JDS],[{C1,C2})
                                    )
                                  )
                               ),
                               Dmds
                            );
          Flag == false -> jcc(  [C1,C2|Clumps],
                                 unsort( compress( sort( JDS ) ) ),
                                 Dmds
                              )
      end;
```

3.3. Sort

I implemented sort using a built-in function from the Erlang Standard Library 1.8.1 "list" module. The lists module contains a function named **sort** whose interface specifies two arguments: a function and the list to be sorted. There were three possible cases which sort had to contend with: a list that contained no elements (empty), one element, or multiple elements. The source for **sort** is provided below.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                        sort.erl
%
%       Case 1:    There is an empty JDS. Return the empty JDS.
%       Case 2:    JDS contains one element. Return the
%                  unmodified JDS since it is already sorted.
%       Case 3:    JDS contains two elements. Compare the
%                  two elements and return them in sorted order.
%       Case 4:    JDS contains multiple elements. Sort them
%                  and use the return value SL and Defs to
%                  return a new JDS.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

sort( [] ) ->
      [];

sort( [ {[{O,N}],Defs} ] ) ->
      [ {[{O,N}],Defs} ];

sort( [ {[{O1,N1}, {O2,N2}],Defs} ] ) ->
      if
            O1 =< O2 -> [ {[{O1,N1}, {O2,N2}],Defs} ];
            O1 > O2 -> [ {[{O2,N2}, {O1,N1}],Defs} ]
      end;

sort( [ {[{O1,N1}, {O2,N2}|T],Defs} ] ) ->
      SL = lists:sort( [{O1,N1}, {O2,N2}|T] ),
      [ {SL,Defs} ].
```

3.4   Compress

The Compress module operates on a JDS that is lexicographically sorted. Compress creates an Entropy Matrix (EM), calculates the maximum compression opportunities, creates a compression definition, replaces the selected data with the compression definition (the compression definition does not appear in the input alphabet), and returns a compressed JDS. Discussion of the Compress module will focus the implementation of the EM data structure since it was one of the more algorithmically complex problems and because of its central importance for calculating compression opportunities.

An EM data structure is a list of 3-tuples in the form of (Position, Count, Clumps). Position refers to the length of match, Count refers to the number of clumps that match, and Clumps is a list of clumps. Thus, an entry of (3,3,Clumps) would be read as: Three clumps match to length three. In order to create an entropy matrix, each data clump in the JDS must be compared to every other clump to establish length-of-match. The first step in this is to establish disjoint sets of matching clumps based on their first two characters. A position counter was initialized to a value of two, and all clumps that had the same two first characters were grouped into what I refer to as a space (note that a clump can match with itself to a length of two, thus creating an entry in the matrix of (2, 1, Clumps)). The clumps in the space were added to the EM and the position counter was incremented by one. Next, the first clump in the space was checked to see that it's length was greater-than-or-equal-to the position counter. If the length was less, the clump was discarded since it would not be able to match with any other clump of greater length. The items in the space were then checked for the next greatest length. Those that matched were added to the EM. Once the space was exhausted, a new space based on the remaining input was created. This process repeated until the entire JDS clump input was consumed.

Once the EM was established, compression opportunities were calculated. By examining the Position and Count elements of each entry in the matrix, it was possible to define the compression opportunity. The compression opportunity was calculated by multiplying the Position element with the Count element. Based on each element in the EM, a new triple

in the form of (Position, Opportunity, Clumps) was created. Thus, a new data structure, an Enhanced Entropy Matrix (EEM) containing triples in the form of (Position, Opportunity, Clumps), was instantiated.

This JC model was designed to pick the highest compression opportunity. The EEM was sorted by descending order based on the Opportunity element of each triple and the highest value was added to a compression data list (CDL). The CDL was designed as a temporary data structure to contain maximum compression opportunities and their associated information (i.e. an entry from the EEM). Multiple maximum values were allowable. Compress then referenced the first entry in the CDL, and extracted the list of clumps from the triple. Each clump found in the list was searched for in the JDS and when found, replaced with a compression symbol. This process continued until the CDL was empty.

The implementation of the Compress module generated the most amount of code. Due to space limitations, I have included the body of Compress, and the following functions: the function build_matrix (control module for creating spaces and the entries in an EM), determine_space (a helper function), build_space (the function that actually builds a space), and create_entry (creates an entry in the EM).

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%                            sort.erl (partial file)
%
%    Case 1:   JDS is empty so return the empty JDS.
%    Case 2:   JDS contains one element. No compression opportunity
%              will be available. Return the unmodified JDS.
%    Case 3:   JDS contains multiple elements. Create an entropy
%              matrix EM. Enhance the EM, compress the JDS and
%              construct a new JDS with compressed data and
%              updated definitions.
%
%    Notes:TMatrix       - is an EM with minimum savings of 4 bytes.
%                 EEM    - an enhanced entropy matrix that contains computed
%                          compression opportunities.
%                 CD     - data structure in the form of a 2-tuple containing
%                          the compression definitions and compressed data.
%                                 Element 1: compression definitions
%                                 Element 2: compressed data.
%                 Data   - the compressed data
%                 NewDefs       - new definitions created during the current
%                                 compression cycle.
%
%                 clean_matrix: returns a matrix where all entries have
%                                      at least two matching clumps.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

-module( compress ).
-export( [compress/1] ).

compress( [] ) ->
         [];

compress( [ { [{O,N}], Defs } ] ) ->
         [ { [{O,N}], Defs } ];

compress([ { [{O,N}|T], Defs } ] ) ->
       EM = build_matrix( [{O,N}|T], [] ),

       TMatrix = if
                             EM /= [] -> clean_matrix( hd(hd( EM ) ), [] );
                             EM == [] -> []
                      end,

       EEM = enhance_matrix( TMatrix, [] );
       CD = compress_it( [{O,N}|T], EEM ),
```

```erlang
          Data = element( 2, CD ),
          NewDefs = element( 1, CD ),

          [ { Data, lists:append( Defs, NewDefs ) } ] ].
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%       Case 1:         The input list is empty (all the data portion of the JDS is
%                       processed). Return the EM.
%       Case 2:         The JDS contains unprocessed data. Process the data and
%                       continue building entries in the EM.
%       Notes:Length  - length of first clump in space
%             Count   - number of clumps in a space
%             Space   - a list of clumps whose first two characters match
%             Matrix  - a temporary matrix containing entries only for
%                           the current space.
%             TrimList a new list from which is the input list minus the
%                           clumps found in the space.            -
%
%             determine_count:       returns a count of clumps whose first
%                                    two characters match.
%             length:                return the length of the first clump
%
%       Note that the build_matrix algorithms operates by creating a space
%       and a local entropy matrix for the space. Once the clumps in the
%       space are processed, the elements in the space are removed from
%       the input list. The temporary Matrix is added to the EM and
%       build_matrix operates on the remaining input list until it is
%       consumed. When the input list is consumed, a complete EM is
%       returned.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```erlang
build_matrix( [], EM ) ->
       EM;

build_matrix( [{O,N}|T], EM ) ->
       Length = length( O ),
       Count = determine_count( Length, [{O,N}|T] ),
       Space = determine_space( [{O,N}|T], Count ),
       Matrix = determine_matrix( Count, Space ),
       TrimList = cut( [{O,N}|T], Count ),
       build_matrix( TrimList, add( Matrix, EM ) ).
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%       Case 1:         If the Count is one, there are no clumps that match for
%                       a minimum length of two. Return an empty space.
%       Case 2:         If the Count is greater than one, construct a space based
%                       on the input list.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```erlang
determine_space( [{O,N}|T], Count ) ->
       if
             Count == 1 -> [];
             Count > 1 -> build_space( [{O,N}|T], Count, [] )
       end.
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%       Case 1:     Count is zero. It is not important as to what is in the
%                   input list since all clumps matching to a length of two
%                   have been removed from the first parameter and placed in
%                   List.
%       Case 2:     The input list has one element left. Add the element
%                   to NewList (which is the current working space). Invoke
%                   build_space with the Count decremented by one.
%       Case 3:     The input list has multiple elements. Add the first element
%                   to NewList (which is the current working space). Invoke
%                   build_space with the Count decremented by one and with the
%                   remainder of the input list.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```erlang
build_space( _, 0, List ) ->
       List;
```

```
build_space( [{0,N}], Count, List ) ->
        NewList = add( {0,N}, List ),
        build_space( [], Count - 1, NewList ).

build_space( [{0,N}|T], Count, List ) ->
        NewList = add( {0,N}, List ),
        build_space( T, Count - 1, NewList ).

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%       Case 1:         The space is empty. Return a matrix entry.
%       Case 2:         The space contains one element. Use find_char to find the
%                       character in the clump 0 at location specified by value
%                       Position - 1. MatchChar specified what character the clump
%                       must match for a given Position value. If CurrChar and
%                       MatchChar are equal. Increment the Count and use the
%                       function add to update the data structure holding the list
%                       of clumps (Temp). If the equality match fails, create an
%                       entry in the temporary EM (Data).
%       Case 3:         Operates like case 2, but for multiple elements in the input
%                       space.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

create_entry( [], MatchChar, Count, Position, Data, Temp ) ->
        add( { Position, Count, Temp }, Data );

create_entry( [{0,N}], MatchChar, Count, Position, Data, Temp ) ->
        CurrChar = find_char( 0, Position - 1 ),

        if
                CurrChar == MatchChar -> create_entry( [],
                                                        MatchChar,
                                                        Count + 1,
                                                        Position,
                                                        Data,
                                                        add( {0,N}, Temp )
                                                      );
                CurrChar /= MatchChar -> create_entry(  [],
                                                        CurrChar,
                                                        0,
                                                        Position,
                                                        add( { Position,
                                                               Count,
                                                               Temp
                                                             ),
                                                        Data
                                                          ),
                                                        []
                                                        )
        end;

create_entry( [{0,N}|T], MatchChar, Count, Position, Data, Temp ) ->
        CurrChar = find_char( 0, Position - 1 ),

        if
                CurrChar == MatchChar -> create_entry( T,

                                                        MatchChar,
                                                        Count + 1,
                                                        Position,
                                                        Data,
                                                        add( {0,N}, Temp )
                                                      );
                CurrChar /= MatchChar -> create_entry( T,

                                                        CurrChar,
                                                        0,
                                                        Position,
                                                        add( { Position,
                                                               Count,
                                                               Temp
                                                             },
                                                             Data
                                                          ),
                                                        []
                                                        )
        end.
```

4. EXPERIMENTAL RESULTS

The JC system testing experimentally investigated the issue of optimal clump generation. Optimal clump generation is important because it directly affects the creation of the JDS, from which the EM is created and used to calculate compression opportunities.

For testing purposes, input data streams for the JC were of finite length and contained a distinct number of characters. The distinct characters had an equal probability of occurrence. In the first experiment three independent variables were chosen: file size (FS), distinct number of characters (DC), and the probability of making a newly encountered character a key (PK). Three different sizes for file size were chosen, four numbers of distinct characters, and five probabilities. The three variables are independent sets with the following members.

$$FS \in \{ 200, 3000, 5000 \}$$
$$DC \in \{ 25, 50, 100, 200 \}$$
$$PB \in \{ 0.0, 0.1, 0.5, 0.9, 1.0 \}$$

A total of 60 (3 x 4 x 5) cases exist and all cases were tested. Data was collected and a model was created using linear, multiple regression. The model has three parameters SIZE (size of file), CHARS (distinct number of characters), and PROB (probability of making a newly encountered character a hash key). Equation 1 summarizes the model.

$$NumOfClumps = 0.5 \times SIZE + 0.1 \times CHARS + 2649.8 \times PROB - 1303.3$$

The model has a correlation coefficient $r^2 = 0.78$ and an F-value = 66.31. The degrees-of-freedom are $v_1 = 3$ and $v_2 = 56$. At a significance level of a = 0.10 the critical F-value is 2.18 [Kac82]. Since the F-value for the model exceeds the critical value, the model is sound and the observed behavior is not due to chance. Table 2 provides a summary for the standard error associated with each independent variable. Standard error1, standard error2, and standard error represent the independent variables SIZE, CHAR and PROB, respectively. I have provided them for verification of the values summarized in Table 3. Table 3 summarizes the significance of the model's independent variables. Significance is determined by comparing the observed t-value of each independent variable with the critical t-value for a t-distribution at a = 0.10. After examining the observed t-value for each independent variable, I found that CHARS was not significant and can be removed from the model [Kac86].

Standard error1	Standard error2	Standard Error
0.0528	0.0655	256.7

Table 2. Standard Error Summary

Variable	t-observed value	Critical t-value	Significant
SIZE	9.602	1.671	YES
CHARS	0.042	1.671	NO
PROB	10.498	1.671	YES

Table 3. Significance of Variables

5. CONCLUSIONS

After completing 75% of the design it was realized that the no-harm-add policy was ineffective for handling certain input sequences. Its inability to process identical, repeated clumps caused the JCC to terminate execution and halt the JC system. This is an unacceptable situation, but the problem has been solved. In the next JC model, a time-stamp will be used to mark clumps that are added to JDS. Thus, the only limitation on JDS will be a memory constraint. This change will be incorporated into the prototype that will be developed for the U.S. Navy in September.

The Pack module, which is responsible for separating control sequences and data, was not implemented in this model of the JC system. JCC's output was a stream of compressed data mixed with control sequences. Pack is being designed and will be incorporated into the next prototype.

The U.S. Navy has specified that the deliverable JC system be coded in Java. Erlang facilitated the development of concise, compact functions that are ideal for formal specification. Aside from the no-harm-add policy, all of the Erlang functions will be used as formal specifications to create the Java JC system prototype.

6. REFERENCES

AVWW96   Joe Armstrong, Robert Virding, Claes Wilkström, Mike Williams. *Concurrent Programming in ERLANG.* Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 1996.

BV99   Harold Abelson, Gerald Jay Sussman, and Julie Sussman. *Structure and Interpretation of Computer Programs.* MIT Press, Cambridge, MA, 2nd edition, 1992.

CRW87   John G. Clearly, Radford M. Neal, and Ian H. Witten. Arithmetic Coding for Data Compression. In *Communications of the ACM*, pages 520-540, June 1987.

Kac82   Sam Kash Kachigan. *Multivariate Statistical Analysis.* Radius Press: New York. 1982.

Kac86   Sam Kash Kachigan. *Statistical Analysis.* Radius Press: New York. 1986.

WY99   John B. Walker and T.C. Yih. *UTA/NAWCAD Joint Development of the Novel Jewel Compression Algorithms with Defense-Relevant Applications.* August 1999.

A Solenoid Valve Model for Pulse Detonation Applications

David S. Jensen
Mentor: Dr. Frank K. Lu
Department of Mechanical and Aerospace Engineering

ABSTRACT

The use of solenoid valves in pulse detonation fuel injection systems requires precise timing and control—the more rapid the response time the more controllable the injection cycle. By applying an excitation voltage to the solenoid, a stronger magnetic field is generated (and at a faster rate), which results in a decrease in response time. This optimization process can be simplified significantly through numerical simulation. This requires the development of a mathematical model of solenoid behavior. Using classical methods, a mathematical model is developed. Some experimentation is necessary in order to characterize the non-linear material reluctance which plays a significant role in the optimization model. The model is also heavily dependent on the effects of flux leakage and core-losses. Expressions are derived to characterize the leakage inductance and induced coil voltage. The determination of these two variables allows for the calculation of the magnetizing current and core-loss resistance. Experimental data should be diligently compared with analytical results, as some scale factor adjustments may be necessary to accurately model a particular solenoid. The model makes use of a simple circuit that is used in conjunction with a large excitation voltage to minimize the response time. The optimal values of the circuit components can then be determined through numerical model simulations. This generalized experimental and analytical procedure can be adapted to a wide class of solenoid valves suitable for pulse detonation applications.

INTRODUCTION

Pulse detonation applications require extremely precise timing and control of the fuel delivery system. As such, electronic fuel injection holds greater developmental opportunity than mechanical systems, which often contain a significant number of moving parts, carry considerable weight, and are difficult to modify once in place. Research into the use of solenoid valves for pulse detonation fuel injection systems has been restricted due to limitations regarding fuel compatibility and insufficient cycle rates, and little has been published on the subject. A recent renewal of interest in this subject can be attributed to a potentially dramatic decrease in response time when a large excitation voltage is applied to a solenoid valve.

This paper develops a general mathematical model of the dynamic behavior of small, high-speed solenoid valves and proposes a mathematical extension to simulate the effects of a controlled excitation voltage. An experimental method for obtaining the data required by the model is also described.

1.0   Overview of a Non-Linear Solenoid

The complete characterization of the dynamic behavior of a solenoid valve requires the modeling of: 1) a non-linear inductance, 2) core loss contributions from eddy currents which create hysteresis in the magnetization curve, and 3) flux leakages [1, 84-94]. These properties can be characterized as components in an equivalent circuit diagram. *Figure 1* illustrates the complete equivalent circuit for a solenoid containing a ferromagnetic core. The variable resistance R_c accounts for eddy current losses and introduces hysteresis in the magnetization curve; the larger its value, the wider the hysteresis loop about the steady state magnetization curve. The inductor L_l accounts for flux leakages and tends to shift the flux curve to the left or right of steady state values. These elements exhibit non-constant values

and are difficult to model, as they are not measurable directly. As such, their effects are often approximated by systematically adjusting their values such that the analytical expressions that contain them reasonably fit the experimental data [2, 108-109].
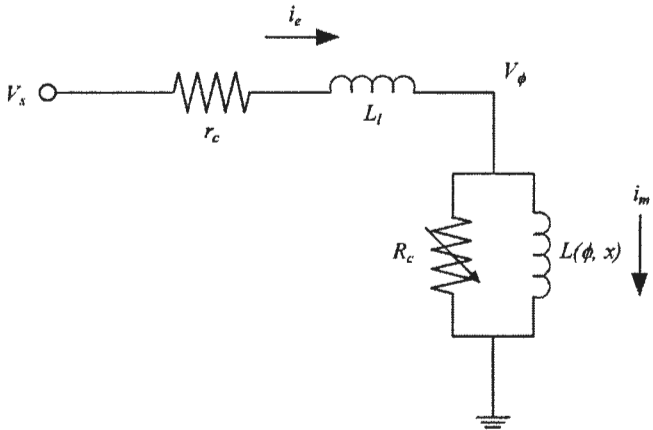
Figure 1. Equivalent Circuit for Non-linear Solenoid

Research has been conducted into minimizing the effects of core-losses and flux leakages by varying core materials and experimenting with their geometry as a means of decreasing response time while increasing efficiency. However, the decrease in response time created by a large excitation voltage has been shown to far exceed decreases produced by other methods [3], [4]. As a result, the optimization section of this paper deals only with the application of an excitation voltage as a means of improving solenoid response time.

In developing the solenoid model, a combination of theory and experimental procedures are presented in order to characterize each component in the equivalent circuit. Although somewhat specific to a particular solenoid geometry, the overall method can be easily adapted to a wide class of solenoids. The following sections analyze each required element or characteristic.

2.0 Development of Solenoid Reluctance

Theoretically, if the material permeability can be assumed constant, the reluctance is equal to the length of the path in the direction of the magnetic flux divided by the product of the permeability and the effective area normal to the path [5, 245-247], [6, 19-23]. For most solenoids, three reluctances must be considered: 1) the fixed air gap reluctance, 2) the variable core air gap reluctance, and 3) the core material reluctance. Since the first two deal with air (whose permeability can be assumed constant), the standard equation can be used. The reluctance of the fixed air gap is then:

$$\mathfrak{R}_g = \frac{g_f}{\mu_o A_g} \tag{1}$$

and the reluctance of the variable core air gap is:

$$\mathfrak{R}_x = \frac{x}{\mu_o A_x} \tag{2}$$

For a cylindrical core geometry, the effective radius r_e (used in both air gap area calculations) is determined by:

$$r_e = \frac{R_2 - R_1}{\ln\left(R_2 / R_1\right)}$$

(3)

Where cylindrical cores are used, the fixed air gap area term is equivalent to the area of a cylinder with effective radius r_e and a length equivalent to the length through which the flux can leak (usually the length of the movable core). For the variable air gap, this is the area of a circle with effective radius r_e.

Evaluating the core reluctance requires a combination of theory and experiment. This is due to the fact that material permeability is not constant for ferromagnetic cores, and thus the standard equation for reluctance cannot be used. Further, ferromagnetic core materials have much larger permeabilities than free space and demonstrate a non-linear relationship between flux density and magnetic field intensity or equivalently, flux and current [7, 165-167]. All ferromagnetic materials possess a flux saturation limit that is characterized mathematically by a horizontal asymptote on the magnetization curve. Fröelich's equation provides a relatively simple analytical method for approximating values on the magnetization curve and is particularly well suited for ranges at or above the knee [6, 63-67]. This equation is likewise easier to manipulate in this model than other approximation expressions and is given below.

$$B = \frac{k_1 H}{H + k_2}$$

(4)

The value of Fröelich's equation is seen in the relationship between B and H (Figure 2). The constant k_1 is equal to the value of flux density at saturation as the field intensity approaches infinity. Further, the constant k_2 is equal to the absolute value of the field intensity as the flux density decreases toward infinity (ignoring any physical significance) [6, 64-65]. In the absence of an actual magnetization curve for a specific material, these two constants must be determined experimentally. It must be realized that due to hysteresis in the magnetization curve, one set of constants describes the opening phase, while a separate set describes the closing phase. Therefore, two sets of "constants" must be obtained.
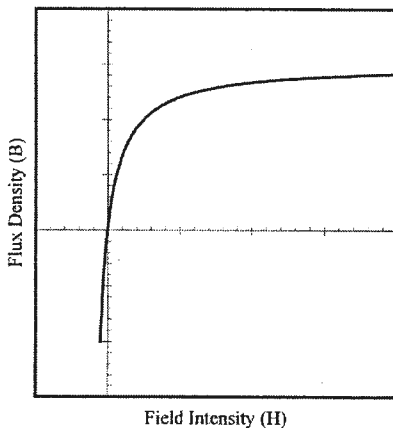
Field Intensity (H)
Figure 2. Theoretical Schematic of Fröelich's Equations

In order to identify the values of these constants, an expression must be derived that makes use of more readily measurable quantities. Current and electromagnetic force represent two quantities that are often easy to obtain experimentally. However, before developing the experiment, Fröelich's equation must be converted to an expression of reluctance. $H = \phi \cdot \mathfrak{R} \big/ \ell$ provides the first relationship, while $B = \phi \big/ A$ provides the second [1, 672-673], [8, 268-269]. Recognizing that these parameters represent quantities pertaining only to the core material, substitutions can be made for both B and H in Fröelich's equation resulting in an expression of material reluctance in terms of flux.

$$\mathfrak{R}_m = \frac{k_2 \ell}{k_1 A_m - \phi} \tag{5}$$

Although the significance of ℓ and A_m is apparent, their presence does not assist in the determination of Fröelich's material constants, and they serve only as scale factors. In order to simplify the identification of these two products, they are combined and referred to as f_1 and f_2, and correspond to $k_1 A_m$ and $k_2 \ell$ respectively [2, 81]. The equation for material reluctance now becomes:

$$\mathfrak{R}_m = \frac{f_2}{f_1 - \phi} \tag{6}$$

From the definition of magnetomotive force, the magnetizing current in a simple coil can be expressed in terms of the product of flux and total reluctance divided by the number of coil turns.

$$i_m = \frac{\phi}{N} \cdot \left(\frac{g_f}{\mu_o A_g} + \frac{x}{\mu_o A_x} + \frac{f_2}{f_1 - \phi} \right) \tag{7}$$

The magnetizing current i_m minus the current contributions from the two air gaps leaves only the current component that produces the flux through the core. Equation (8) characterizes magnetomotive force in terms of core material reluctance and is used to calculate the values of the two material constants.

$$\frac{N}{\phi} \cdot i_m - \left(\mathfrak{R}_g + \mathfrak{R}_x \right) = \frac{f_2}{f_1 - \phi} \tag{8}$$

The current component i_m is not directly measurable; however, prior research has demonstrated that through the central portion of the cycle, the magnetizing current varies little from the exciting current, and measured values of the exciting current taken from reasonably spaced points at or above the knee are sufficient for the determination of Fröelich's constants [2, 115], [9, 425]. It should likewise be expected that the core-loss current would be small for an efficient solenoid design. The core-loss current effects can be further explained through the analogy that the magnetizing current lags the exciting current initially and finally during each cycle, for it is during these two periods that the greatest amount of current flows through the core-loss resistor, creating the most pronounced hysteresis effect. As a caution, it cannot be assumed that every solenoid is designed to optimal specifications. For example, if the number of coil windings is increased beyond the number that produces flux saturation, the only result is increased energy consumption—without a corresponding

increase in solenoid force. From the equivalent circuit perspective, a larger portion of the exciting current runs through the core-loss pathway and is dissipated as heat energy. Therefore, before using excitation current values in the determination of Fröelich's constants, an experiment should be conducted to assess the impact of current variation on the constant values.

First, the solenoid core should be fixed in both the maximum and minimum air gap positions while applying a constant excitation voltage such that the excitation current can be measured across a resistance with respect to time. The variation between these two current profiles serves as an indicator of the efficiency of the solenoid design. For an efficient design, the parameters f_1 and f_2 can be evaluated using two reasonably spaced excitation current values taken from above the knee. Then using the same flux value, the two constants should be determined using a variety of current values less than the excitation current values. In this manner the effect of reduced magnetizing current values on the two constants can be examined. It is generally true that for most solenoid designs, the use of excitation current values in the determination of Fröelich's constants is acceptable. Again this experiment must be conducted separately for the opening and closing phases.

A method for computing flux must next be derived and is found in the experimental measurement of electromagnetic force, which offers a pathway to its calculation. By definition, the electromagnetic force can be obtained from the negative of the partial differential of the stored energy with respect to the variable air gap as illustrated by equation (9) [1, 683], [7, 167-168], [9, 421].

$$F_e(i_e, x) = -\frac{\partial w(i_e, x)}{\partial x} = -\frac{\partial}{\partial x}\left(\int_0^{\phi} Ni_e \cdot d\phi\right) = -\frac{\partial}{\partial x}\left(\int_0^{\phi} \Re\phi \cdot d\phi\right) \quad (9)$$

In reviewing each of the three reluctance components previously obtained, it is seen that only the variable core air gap reluctance component contains the variable x, and equation (9) simplifies to:

$$F_e(i, x) = -\frac{\partial}{\partial x}\left(\int_0^{\phi} \frac{x\phi}{\mu_o A_x} d\phi\right) = -\frac{\phi^2}{2\mu_o A_x} \quad (10)$$

Section 2.1 describes an experimental method for obtaining electromagnetic force from which the flux can be calculated (it should be noted that the negative sign indicates the proper direction of the force but is not needed in the flux dependent current equation, and the absolute values should be used).

From the force and current data collected experimentally, a graph of flux versus current is made (only the data that resides at or above the knee of the current vs. time curve should be used). The data from two reasonably spaced points are then substituted into equation (8), and the resulting two equations are solved simultaneously for the unknown constants. The pair of constants to be used for closing phase calculations is obtained in the same manner. Once the constants have been determined, the variable material reluctance can be calculated (for both opening and closing phases) from the corresponding flux values. Graphically, one can also read off the values of material reluctance as the slope at a point on the graph of magnetomotive force versus flux [6, 22].

2.1   General Force Equation and Experimental Procedures

As stated in the previous section, the electromagnetic force data must be obtained experimentally. Naturally the technique employed to measure this force depends largely on the solenoid valve geometry. The fuel flow requirements of a pulse detonation engine are best suited to a valve type that can be modeled as a simple, single spring-return valve as illustrated in *Figure 3*. Such a design facilitates the flow of gaseous fuels, allows for a precise opening and closing phase, and is responsive to optimization.
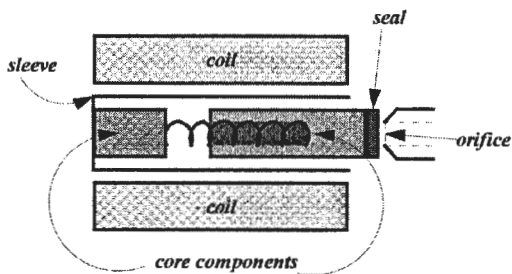
Figure 3. Sample Solenoid Internal Geometry

Additionally, this type of solenoid valve can be modeled as a damped mass/spring system as illustrated in equation 11.

$$F_e(i_e, x) + F_g + p + K(x_{max} - x) = m\frac{d^2x}{dt^2} + C_d\frac{dx}{dt} \qquad (11)$$

Since most manifold type solenoids operate in the horizontal position, the gravitational force is neglected. Solving for the electromagnetic force yields:

$$F_e(i_e, x) = m\frac{d^2x}{dt^2} + C_d\frac{dx}{dt} - F_g - K(x_{max} - x) + p \qquad (12)$$

Special care should be taken in evaluating equation (12) due to the vector quantities involved so that the signs indicate the proper directions of the various forces. In designs such as that illustrated by *Figure 3*, the effect of the gas pressure force becomes essentially zero as soon as the seal moves away from the orifice. This effect can be achieved more rapidly by the addition of a small vent hole in the movable core component, thus equalizing the gas pressure throughout the interior of the valve. With this modification, the gas pressure force can be neglected in equation (12). As a result of this benefit, it is suitable to conduct initial optimization experiments without the complications of added gas flow.

Not only must a suitable valve possess a reasonable geometry, it must be capable of satisfying other demands of pulse detonation: 1) a sufficient flow coefficient, 2) a fast response time, and 3) high frequency operation capabilities. A search of production valves indicated the Numatech® "TM" series valve as a likely candidate. With only minor adjustments this valve can be adequately modeled as a simple spring-return valve. The following experimental description applies specifically to the Numatech® valve, but the general procedures can be adapted to a variety of valve types.

By separating the solenoid and core components from the rest of the valve and preloading it in a brace to the original configuration, the dynamic response can be observed independent of the valve casing. Further, a laser assembly can be used to pass a beam of light perpendicular to the exposed movable core. Once the solenoid is energized (or de-energized), a photo-detector mounted on the opposite side measures the changing light intensity versus time as the core moves to the final position. The beam should be confined such that when the core is in the closed position, the photo-detector reads at or near zero. Since it is the response time that is of interest, each experiment should be conducted for a single application or removal of an excitation voltage. The photo-detector is connected to a digital oscilloscope that acquires and stores the data. A curve is then fitted to the data representing a function of core displacement in terms of time. The resulting function coupled with the corresponding data from the first and second derivatives of this function can then be used in equation (12) to calculate the electromagnetic force. The force results are then used to calculate the flux given by equation (10). It is important to remember that separate experiments must be conducted for both the opening and closing phases.

The damping coefficient C_d cannot be readily determined but can be approximated by numerical simulation such that the value of the flux dependent, induced coil voltage is slightly less than the source voltage minus the drop across the pure coil resistance.

2.2   Induced Coil Voltage

According to Faraday's law, the induced coil voltage (which occurs as a result of the changing magnetic flux) is equal to the product of the number of coil turns and the derivative of the flux with respect to time [6, 76-77]. In the previous section, the flux was calculated from the force data generated by equation (12). Fitting a curve to the flux allows a convenient solution to the induced coil voltage which, according to Faraday's law is:

$$V_\phi = N\frac{d\phi}{dt} \tag{13}$$

Alternatively, flux is a function of both the magnetizing current and the variable air gap length, and the full differential of flux with respect to time yields an alternate equation for the induced coil voltage.

$$V_\phi = N\cdot\frac{\partial\phi}{\partial i_m}\frac{di_m}{dt} + N\cdot\frac{\partial\phi}{\partial x}\frac{dx}{dt} \tag{14}$$

Again, differentiating equation (7) yields an equation in terms of $d\phi$, di_m, and d_x:

$$N \cdot di_m = \left[\frac{g_f}{\mu_o A_g} + \frac{x}{\mu_o A_x} + \frac{f_1 f_2}{\phi^2 \left(\dfrac{f_1}{\phi} - 1 \right)^2} \right] d\phi + \frac{\phi}{\mu_o A_x} dx \tag{15}$$

Simplifying equation (15) and solving for the time rate of change of flux yields:

$$\frac{d\phi}{dt} = \left[\frac{N}{\Re_g + \Re_x + \dfrac{f_1 f_2}{(f_1 - \phi)^2}} \right] \frac{di_m}{dt} - \left[\frac{\phi / \mu_o A_x}{\Re_g + \Re_x + \dfrac{f_1 f_2}{(f_1 - \phi)^2}} \right] \frac{dx}{dt} \tag{16}$$

The coefficient of the di_m/dt term in equation (16) is equal to the partial differential of flux with respect to magnetizing current as given in equation (14), while the corresponding coefficient of the dx/dt term in equation (16) is equal to the partial differential of flux with respect to the variable core air gap length, also from equation (14). Substituting these two terms into equation (14) completes the alternate expression of the induced coil voltage.

$$V_\phi = \left[\frac{N^2}{\Re_g + \Re_x + \dfrac{f_1 f_2}{(f_1 - \phi)^2}} \right] \frac{di_m}{dt} - \left[\frac{N\phi / \mu_o A_x}{\Re_g + \Re_x + \dfrac{f_1 f_2}{(f_1 - \phi)^2}} \right] \frac{dx}{dt} \tag{17}$$

2.3   Flux Leakage Inductance

From the equivalent circuit diagram (*Figure 1*), the flux leakage inductance can be determined by applying Kirchhoff's voltage law. Substituting equation (13) for the induced coil voltage and simplifying yields an expression for the flux leakage inductance.

$$L_\ell = \frac{V_s - i_e r_c - N \dfrac{d\phi}{dt}}{\dfrac{di_e}{dt}} \tag{18}$$

From the experimental data, curves can be fit to both flux and exciting current such that values of the flux leakage inductance are easily obtained by numerical simulation.

2.4   Non-Linear Inductance

From the definition of inductance, the coefficient of the di_m/dt term in equation (17) describes the non-linear coil inductance $L(\phi, x)$.

$$L(\phi,x)=\frac{N^2}{\mathfrak{R}_g + \mathfrak{R}_x + \dfrac{f_1 f_2}{(f_1-\phi)^2}} \tag{19}$$

Similarly, the coefficient of the dx/dt term of the same equation is proportional to the non-linear coil inductance by the factor $\phi/(N\mu_o A_x)$ [2, 83]. Taken separately, each complete term represents a voltage contribution that is a direct result of the rate of current in the coil and the core velocity, respectively.

2.5   Equivalent Circuit Analysis

Figure 1 is reproduced as *Figure 4* with the addition of a box representing an optimization circuit. The appearance of eddy currents and hysteresis in the magnetization curve produce a heat-related power loss that can be modeled as the variable resistance R_c in parallel with the solenoid coil [1, 693-694], [6, 88-90]. It is worthwhile to note that the
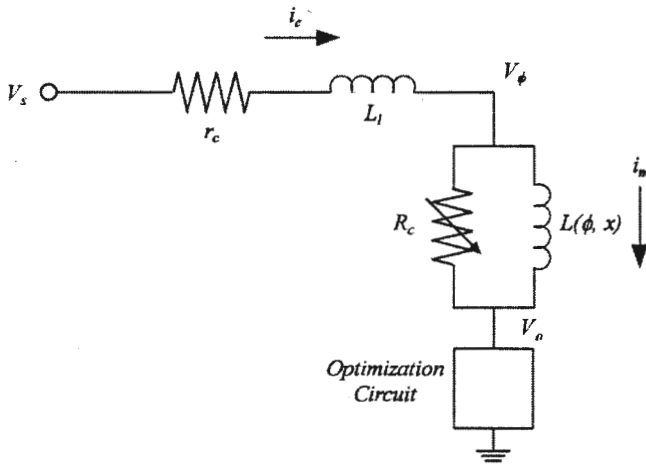
Figure 4. Equivalent Circuit for Optimized Non-linear Solenoid

eddy current pathways become less dense toward the center of the core creating a nonuniform flux distribution over the core's cross section [6, 88]. From the diagram, the equation for V_s can be written as:

$$V_s = i_e r_c + L_\ell \frac{di_e}{dt} + V_\phi \tag{20}$$

Substituting $V_\phi - V_o = R_c(i_e - i_m)$ into equation (20) yields an alternate description of V_s.

$$V_s = i_e r_c + L_\ell \frac{di_e}{dt} + R_c\left(i_e - i_m\right) + V_o \tag{21}$$

Knowing V_ϕ from Faraday's Law allows the determination of the magnetizing current from equation (17), or more directly, equation (7). Substituting the resulting expression of magnetizing current into equation (21) allows for the complete characterization of the core-loss resistance R_c.

As a final expression, $V_\phi - V_o = R_c\left(i_e - i_m\right)$ is substituted into equation (17); simplifying, solving for di_m/dt, and integrating the result yields:

$$i_m = \int\left\{\frac{R_c\left(i_e - i_m\right)}{N^2}\left[\mathfrak{R}_g + \mathfrak{R}_x + \frac{f_1 f_2}{\left(f_1 - \phi\right)^2}\right] + \frac{\phi}{\mu_o A_x N}\frac{dx}{dt}\right\}dt \tag{22}$$

3.0 Optimization Circuit Analysis

Figure 5 illustrates an optimization circuit concept used successfully in controlling excited solenoid valves in the automotive industry [2, 105]. The capacitor is selected in order to reduce the voltage across the solenoid coil quickly to the nominal value. The zener diode operates in the breakdown range and is selected such that its breakdown voltage is equal to the nominal solenoid voltage and its current characteristic equal to the nominal current. The resistor facilitates the rapid discharge of the capacitor [9, 417-418].
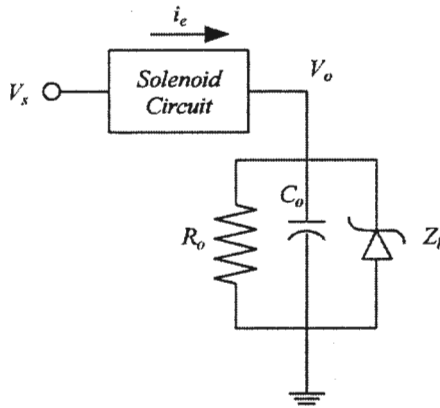
Figure 5. Solenoid Optimization Circuit

Applying Kirchhoff's current law to the node at V_o results in the following expression that is valid only when V_o is less than the breakdown voltage of the zener diode. For all other values, V_o is equal to the breakdown voltage.

$$\frac{dV_o}{dt} + \frac{1}{R_o C_o}V_o = \frac{i_e}{C_o} \tag{23}$$

In addition, the following previously derived equations can be adapted to complete the optimization circuit model.

$$V_\phi - V_o = \left[\dfrac{N^2}{\mathfrak{R}_g + \mathfrak{R}_x + \dfrac{f_1 f_2}{(f_1 - \phi)^2}} \right] \dfrac{di_m}{dt} - \left[\dfrac{N\phi / \mu_o A_x}{\mathfrak{R}_g + \mathfrak{R}_x + \dfrac{f_1 f_2}{(f_1 - \phi)^2}} \right] \dfrac{dx}{dt} \qquad (24)$$

$$\dfrac{di_e}{dt} = \dfrac{1}{L_\ell} \left[V_s - V_o - i_e r_c + R_c (i_m - i_e) \right] \qquad (25)$$

By using the model as developed, numerical simulations can be conducted to determine appropriate values for the optimization circuit components under a wide variety of excitation voltages. When fully optimized, the solenoid valve response time will decrease under the application of an excitation voltage, and a large current boost will be delivered. The optimization circuit will rapidly reduce the current and voltage to nominal levels, thus preventing the solenoid from overheating. Upon removal of the voltage source, the circuit facilitates the rapid collapse of the magnetic field, thus allowing for a more rapid closing phase. Actual experiments can then be conducted to validate the theoretical results. In previous studies, excitation voltages as high as three times the nominal values have produced significant results without compromising the life of the solenoid valve [2], [9].

CONCLUSIONS

The accurate modeling of a solenoid under the influence of an excitation voltage introduces several significant complications. The first is the non-linear material reluctance which plays a significant role in the model. Unfortunately, material parameters vary significantly from product to product, and no unified approach exists for determining the required constants. Fröelich's equation offers perhaps the most convenient solution and can be completed with some experimentation. Obtaining the correct force values for the required flux calculations is complicated due to the effect of the damping coefficient term in the force equation which must also be accounted for. Although it is the magnetizing current that is required to properly obtain Fröelich's constants, the exciting current values can be used over periods at or above the knee of the current versus time curve, where typically both currents are approximately the same value. This is due to the fact that in the middle section of the curve, core-loss effects have been shown to be essentially zero [9, 425]. As a caution, this method is valid so long as the solenoid design is reasonably efficient. If this characteristic does not hold for a particular solenoid, then the determination of the required magnetizing current values is complicated. It should be remembered that all reluctance components must be included for an accurate representation of solenoid behavior.

Effects due to leakage inductance and core-loss resistance also complicate the model. Since neither of these values is constant, each can be assumed to be a function of some variable. In most research efforts, they are modeled as a linear function of the variable air gap and occasionally, flux [2], [9]. The methods presented in this paper offer a systematic path for deriving expressions. Since each subsequent expression is dependent on the accuracy of the first, it may be necessary to adjust the analytical expressions by various scale factors so that they agree with the experimental data to an acceptable degree of accuracy

[2]. Finite element methods have also been employed in the modeling of excited solenoid valves with good success, so long as the variable air gaps are large (7 mm) [4]. For pulse detonation, however, large air gaps inherently create response times that are not conducive to 100 Hz cycles.

In order to simplify the modeling process so that the optimization process can be expedited, procedures for determining the damping coefficient should be developed. This model is particularly well suited to numerical simulations conducted on curve fit estimations of experimental data.

Future research to be conducted by the author will include experimentation as described in this paper, the verification and modification (as necessary) of the model as presented, and optimization experiments of a solenoid valve under high-pressure gas flow. This will lead to the design and development of an electronic fuel injection system for a practical pulse detonation engine.

LIST OF FIGURES AND NOMENCLATURE

LIST OF FIGURES

Number Description

NOMENCLATURE

Symbol Description

A_g - fixed air gap cross-sectional area normal to the magnetic flux path (m^2)

A_m - core cross-sectional area normal to the magnetic flux path (m^2)

A_x - variable air gap cross-sectional area normal to magnetic flux path (m^2)

B - magnetic flux density (Wb/m^2)

C_d - damping coefficient (kg/s)

C_o - optimization capacitance (F)

$F_e(i_e, x)$ - electromagnetic force of attraction (N)

F_g - gas pressure force exerted on the core in the direction of x (N)

F_m - magnetomotive force (A-T)

f_1 - material reluctance constant (Wb)

f_2 - material reluctance constant (A)

g_f - fixed air gap length in the direction of the magnetic flux path $[R_2 - R_1]$ (m)

H - magnetic field intensity (A/m)

i_c - core flux component of current (A)

i_e - total or excitation current (A)

i_m — magnetization current (A)

K — spring constant over linear range equal to ^{x} (N/m)

k_1 — flux density saturation value (Wb/m^2)

k_2 — negative field intensity saturation value (A/m)

L_ℓ — flux leakage inductance (H)

$L(\phi,x)$ — non-linear coil inductance (H)

l — length of core through which magnetic flux traverses (m)

m — mass of movable valve core assembly (kg)

N — number of coil turns

p — primary spring preload compression force (N)

R_1 — radius of the solenoid core (m)

R_2 — sum of the solenoid core and the fixed air gap length (m)

R_c — eddy current loss resistance (Ω)

R_o — optimization resistance (Ω)

r_c — coil resistance (Ω)

r_e — effective radius used in area calculations (m)

V_s — source voltage (V)

V_o — optimization circuit voltage (V)

V_ϕ — induced coil voltage (V)

x — variable air gap length in the direction of the magnetic flux path (m)

x_{max} — maximum variable air gap length (m)

Z_b — breakdown voltage zener diode

$\Re$ — total reluctance (A/Wb)

$\Re_g$ — fixed air gap reluctance (A/Wb)

$\Re_m$ — material or core reluctance (A/Wb)

$\Re_x$ — variable air gap reluctance (A/Wb)

μ_o — permeability of free space ($4\pi\times10^{-7}$ Wb/A·m)

ϕ — magnetic flux (Wb)

ACKNOWLEDGEMENTS

REFERENCES

[1] Hambley, A. R., *Electrical Engineering—Principles and Applications*, Prentice Hall, New Jersey, 1997.

[2] Hong, H., "Optimum Performance of Solenoid Injectors for Direct Injection of Gaseous Fuels in IC Engines," Dissertation. Concordia University, Montreal, Quebec, Canada, 1995.

[3] Ohdachi, Y., Kawase, Y., Murakami, Y., and Inaguma, Y., "Optimum Design of

Dynamic Response in Automotive Solenoid Valve," *IEEE Transactions on Magnetics*, Vol. 27, No. 6, Nov. 1991, pp. 5226-5228.

[4]   Mitsutake, Y. and Hirata, K., "Dynamic Response Analysis of a Linear Solenoid Actuator," *IEEE Transactions on Magnetics*, Vol. 33, No. 2, March, 1997, pp. 1634-1637.

[5]   Kraus, J. D., *Electromagnetics,* McGraw Hill, New York, 1953.

[6]   Hunt, W. T., and Stein, R., *Static Electromagnetic Devices*, Allyn and Bacon, Boston, 1963.

[7]   Coren, R. L., *Basic Engineering Electromagnetics*, Prentice Hall, New Jersey, 1989.

[8]   Demarest, K. R., *Engineering Electromagnetics*, Prentice Hall, New Jersey, 1998.

[9]   Hong, H., Krepec, T., Cheng, R. M. H., "Transient Response of Fast Acting Solenoids in Automotive Applications," *Journal of Circuits, Systems, and Computers,* Vol. 4, No. 4, 1994, pp. 415-428.

The Effects of Strain Gradients on Strain Measurement in a Glass/Vinyl-Ester Laminate

Terrence E. Johnson

Mentor: Dr. John M. Kennedy, Department of Mechanical Engineering, Clemson University

Mentor: Dr. Kendall Harris, Department of Mechanical Engineering, The University of Texas at Arlington

ABSTRACT

The objective of this paper is to evaluate the effect of strain gradients on strain measurement in a glass/vinyl-ester material system. This research is directly involved with the Navy's effort to replace steel with glass/vinyl-ester on the exterior walls of its Aegis Destroyer connected helicopter hangars. This research will allow the Navy to better understand the mechanical properties of the glass/vinyl-ester material system. The effects of strain gradients on strain measurement was studied by varying the size of the strain gages ($\frac{1}{2}$" x $\frac{1}{2}$", 1" x $\frac{1}{4}$", $\frac{1}{16}$" x $\frac{1}{16}$") and the location of a strain gage in five different positions relative to a q-cell on the glass/vinyl-ester material (center of 90∞ fibered q-cell, center of 0∞ fibered q-cell, center of unit cell, center of transverse boundary of two q-cells, center of longitudinal boundary of two q-cells). The aforementioned strain gages of various sizes were bonded to four dog–bone shaped samples of glass/vinyl-ester material. Each sample was put into tensile loading and observed for induced strain. All data was analyzed using charts and Microsoft's t-stat program. It has been concluded that strain gradients have a significant effect on small strain gages when small strain gages are compared to large strain gage. Therefore, large strain gages are recommended for measuring strain in glass/vinyl-ester material.

INTRODUCTION

One of the Navy's research objectives is to demonstrate affordable composite material structures with enhanced war-fighting capability by integrating a composite material helicopter hangar into the Aegis destroyer. Instead of using steel for the hangar structure, glass/vinyl-ester will be used. The benefits of using glass/vinyl-ester in lieu of steel are that it will reduce corrosion, electromagnetic signature, and cost, but, at the same time, expand the industry base for marine composite design and manufacturing technology, and decrease the weight of the hangar by 25% over the existing steel design.

Clemson University is working in collaboration with the Navy to determine the performance of a glass/vinyl-ester material system that will be used to manufacture the hangar. In this study, the mechanical properties of a representative laminate are being determined at service conditions. The laminate consists of layers of plain-weave glass cloth. The width of the glass tow in the weave is approximately $\frac{1}{4}$ inch. Consequently, there may be strain gradients in the plane of the laminate. Masters and Ifju [1] have investigated the effect of strain gradients in textile composites on mechanical properties when strains are measured using strain gages. Their work indicated that sufficient accuracy could be obtained using strain gages that were larger than the unit cell size of the fiber architecture. Unfortunately, most standard material property tests utilize small test samples that may require the use of strain gages that are smaller than the unit cell size. Thus, the size and location of the strain gage relative to the fiber architecture are important considerations when measuring strain to determine material properties. Therefore, the objective of this research was to determine how strain gradients in a glass/vinyl-ester laminate effect strain measurements and thus mechanical properties.

The investigation was conducted by measuring strain at several locations relative to the unit cell with three different strain gages. Elastic stiffness was determined using each strain gage and statistical analysis was performed to determine significant effects of gage size and location.

EXPERIMENTAL PROGRAM

The material tested herein had a stacking sequence that produced a nearly quasi-isotropic laminate. The material was prepared in large panels using the Scrimp process. Four dog-bone shaped tensile coupons were machined from a large panel. The specimen had an overall length of 24 inches and a width of $1\frac{7}{8}$ inches. The width of the specimen was reduced to $1\frac{1}{2}$ inches, 2 inches away from each end with a 3-inch diameter fillet.

Gages of different sizes ($\frac{1}{16}" \times \frac{1}{16}"$, $1" \times \frac{1}{4}"$, $\frac{1}{2}" \times \frac{1}{2}"$) were bonded to each sample. These gages can be observed in Figure 1. The strain gages were bonded to a sample using standard strain gage application techniques. In addition, each gage was attached to five different locations on the sample relative to a q-cell as shown in Figure 2. Each sample was tested in tension in an Instron 56.2 kip Model 1332 servo-hydraulic machine. The test was run in load control and had a loading rate of 600 lbf/min. Data was acquired using LabVIEW and analyzed using Microsoft Excel.

$$\frac{1}{16}" \times \frac{1}{16}" \qquad 1" \times \frac{1}{4}" \qquad \frac{1}{2}" \times \frac{1}{2}"$$

Figure 1: Strain Gage Size

center of q-cell (two directions)

center of unit cell

center of transverse boundary of two q-cells

center of longitudinal boundary of two q-cells
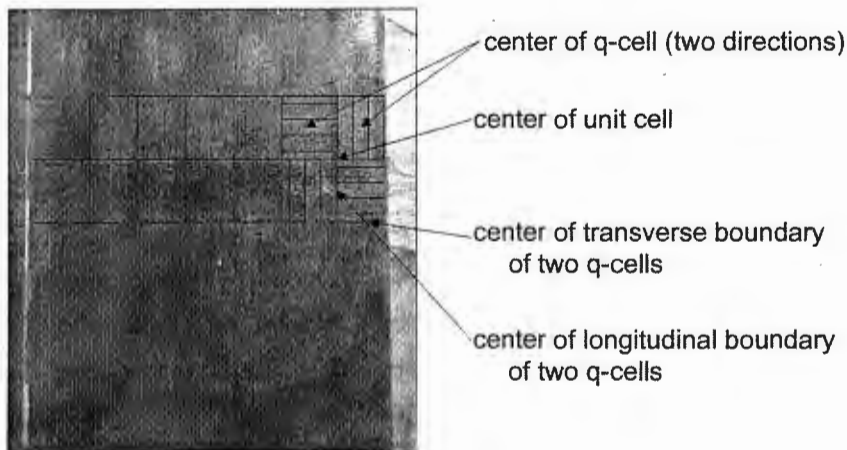
Figure 2: Strain Gage Locations

RESULTS AND DISCUSSION

All test conducted using Instron's servo-hydraulic machine were successful. All glass/vinyl-ester samples tested in the machine were loaded to failure. Failure occurred within the samples at approximately 10,000 to 15,000 pound force. During the test, load, position, and strain data was recorded. Strain was converted into elastic modulus and analyzed empirically using stress-strain curves shown in Figures 3 through 5, below.
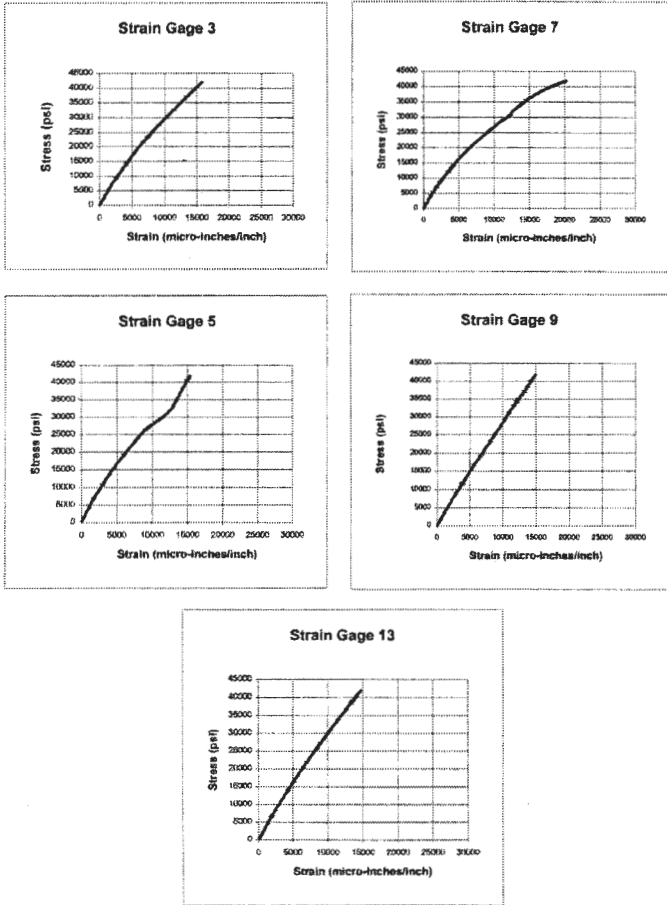
Figure 3: Stress-Strain Curve for the $^1/_{16}$" x $^1/_{16}$" Gage
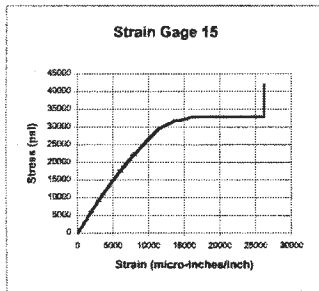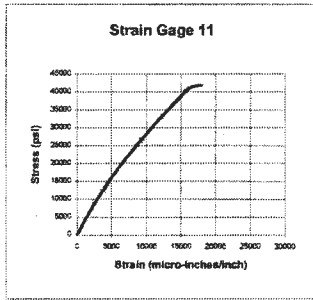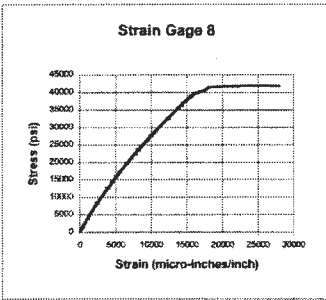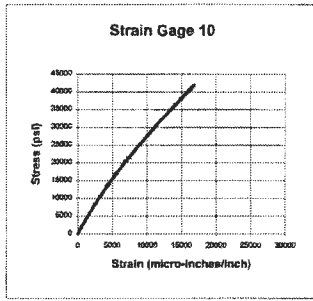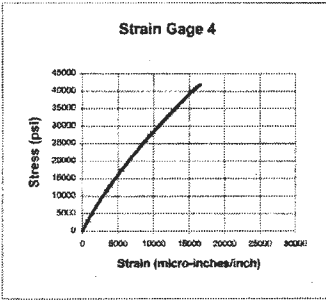
Figure 4: Stress-Strain Curve for the 1" x ¹/₄" Gage
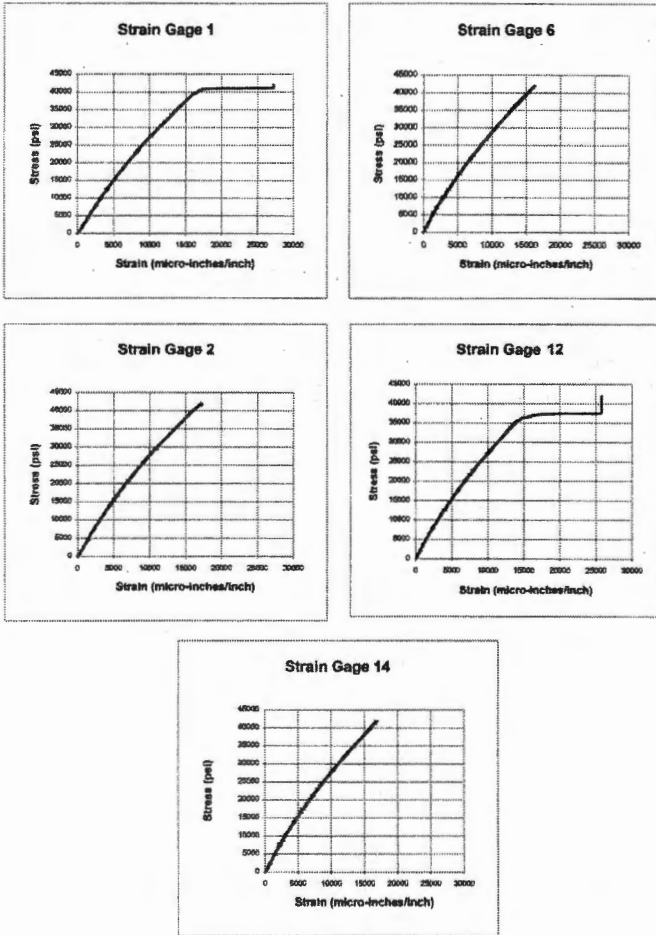
Figure 5: Stress-Strain Curve for the $\frac{1}{2}$" x $\frac{1}{2}$" Gage

From the stress-strain curves, the point at which non-linearity occurred was observed and data was truncated at this point. A linear regression analysis was preformed on truncated data to determine modulus. These values can be observed in Table 1 (next page).

LOCATION	SIZE			Average
	$1/2$" x $1/2$"	$1/16$" x $1/16$"	1" x $1/4$"	
Center of unit cell	3.283551247	3.245828094	3.497374328	Sample 2
	3.436058871	3.245828094	3.509919081	Sample 3
	3.576863457	2.866201407	3.517888362	Sample 4
	3.302375508	3.004237298	3.449309089	Sample 5
Average:	*3.399712271*	*3.090523723*	*3.493622715*	*3.327952903*
Transverse axis	3.504000695	3.584317769	3.319493557	Sample 2
	3.494487280	4.221962892	3.592287688	Sample 3
	3.673339941	3.020765966	3.437549219	Sample 4
	3.543121643	3.585880663	3.526728964	Sample 5
Average:	*3.55373739*	*3.603231823*	*3.469014857*	*3.54199469*
Longitudnal axis	3.229045242	3.597865967	3.445948403	Sample 2
	3.606051542	3.210531411	3.517888362	Sample 3
	3.672414070	3.437549219	3.483433014	Sample 4
	3.510710887	3.962144838	3.682137266	Sample 5
Average:	*3.504555435*	*3.552022859*	*3.532351761*	*3.529643352*
0 degree fibers	3.463830401	3.788772342	3.645064167	Sample 2
	3.385245899	3.547245882	3.45448198	Sample 3
	3.310585880	3.547132506	3.285454732	Sample 4
	3.745739678	4.092835864	3.208314341	Sample 5
Average:	*3.476350465*	*3.743996649*	*3.398328805*	*3.539558639*
90 degree fibers	3.332518039	3.516435519	3.305711945	Sample 2
	3.244673454	3.693042461	3.903799661	Sample 3
	3.378868567	4.205194116	3.555943088	Sample 4
	3.179000769	3.792983054	3.647045648	Sample 5
Average:	*3.283765207*	*3.801913787*	*3.603125086*	*3.562934693*
AVERAGE	3.450284943	3.548188767	3.494962126	

Table 1: Linear Regression Analysis

The values in Table 1 represent modulus values of the strain gage at the aforementioned locations relative to a q-cell. The table is separated into sections that investigate moduli within a single sized strain gage at a selective location. The values located within each section was used in evaluating strain gage measurement when strain gage size and strain gage location relative to a q-cell were held constant. The averages for each gage size at a distant location relative to a q-cell was calculated and used in an empirical analysis of strain. A statistical analysis was conducted using each size-position section, comparing this section to other sections within the table.

In holding strain gage location relative to a q-cell constant, variation in strain is observed across the specimen. Empirically a variation in modulus is observed at the center of a unit location relative to a q-cell. The largest variation in modulus occurs between the 1" x $1/4$" and the $1/16$" x $1/16$" strain gages at the center of a unit location. The $1/2$" x $1/2$" strain gage shows the least amount of variation from the average modulus than the 1" x $1/4$" strain gage and the $1/16$" x $1/16$" strain gages. Statistically, a comparison in modulus between the 1" x $1/4$" and the $1/16$" x $1/16$" and between the $1/2$" x $1/2$" and the $1/16$" x $1/16$" strain gages at the center of a unit location shows a significant amount of deviation between strain measurement within the compared size gages. That is, when bonding strain gages at the center of a

unit cell location, the size of the strain gage affects strain measurement. Therefore, careful attention must be given to strain gage size at this position relative to a q-cell along the material. In choosing a strain gage at the center of a unit cell location, special consideration should be given to the $1/2$" x $1/2$" strain gage because its average modulus measurement is closer to the average modulus measurement (ref. Table 1) than any other size strain. This can be observed in Figure 6.
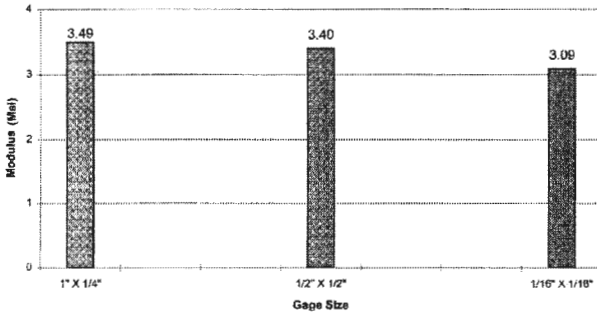

Figure 6: Average Modulus at Center of Unit Cell Location

A significant variation in modulus is not detected between the strain gages at either the transverse axis, the longitudinal axis or the 0° fibers locations relative to a q-cell. Therefore, any size strain gage of the three aforementioned size gages is recommended at the transverse axis, longitudinal axis and 0° fibers locations relative to a q-cell. An empirical evaluation of strain gages at the transverse axis location relative to a q-cell showed large variations in average modulus measurement in the $1/16$" x $1/16$" and the 1" x $1/4$" strain gages. The $1/16$" x $1/16$" and the 1" x $1/4$" strain gages deviated by more than 6% from the average strain gage measurement at the transverse axis location. The smallest amount of variation from the average modulus measurement at the transverse axis location was observed in the $1/2$" x $1/2$" strain gage. This strain gage deviated by 1% from the average modulus measurement at this location. The average modulus recorded at the transverse axis location was 3.54 Msi (Mega-pounds per square inch). The average standard deviation within the strain gage modulus from the average modulus was 0.230. This can be observed in Figure 7.
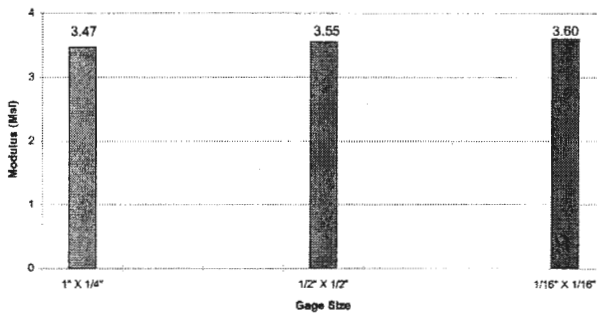
Figure 7: Average Modulus in Transverse Axis Location

An empirical evaluation of strain gages at the longitudinal axis location relative to a q-cell showed large variations in average modulus measurement from the average modulus measurement at the longitudinal axis location in the $1/16$" x $1/16$" and the $1/2$" x $1/2$" strain

gages. The $\frac{1}{16}$" x $\frac{1}{16}$" and the 1" x $\frac{1}{2}$" strain gages deviated by more than 2% from the average strain gage measurement at the longitudinal axis location. The smallest amount of variation from the average modulus measurement at the longitudinal axis location was observed in the 1" x $\frac{1}{4}$" strain gage. The 1" x $\frac{1}{4}$" strain gage deviated by .2% from the average modulus measurement at this location. The average modulus recorded at this location was 3.53 Msi (Mega-pounds per square inch). The average standard deviation calculated at this location was .205. This can be observed in Figure 8.
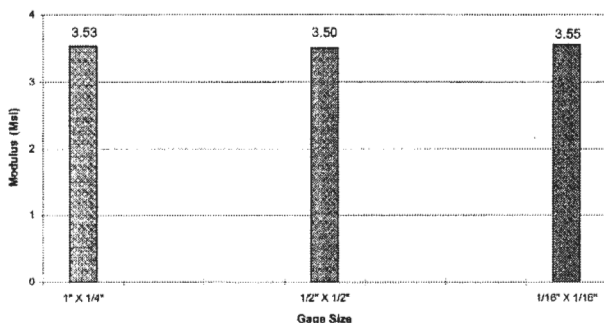
Figure 8: Average Modulus in Longitudinal Axis Location

An empirical evaluation of strain measurement at the 0° fibers location relative to a q-cell showed large variations in modulus from the average modulus measurement at the 0° fibers location in the $\frac{1}{16}$" x $\frac{1}{16}$" and the 1" x $\frac{1}{4}$" strain gages. The $\frac{1}{16}$" x $\frac{1}{16}$" and the 1" x $\frac{1}{2}$" strain gages deviated by more than 14% from the average modulus measurement at the 0° fibers location. The smallest amount of variation from the average modulus measurement at the 0° fibers location was observed in the $\frac{1}{2}$" x $\frac{1}{2}$" strain gage. This strain gage deviated by 6% from the average modulus measurement at this location. The average modulus recorded at this location was 3.54 Msi. The average standard deviation calculated at this location is .214. This can be observed below in Figure 9.
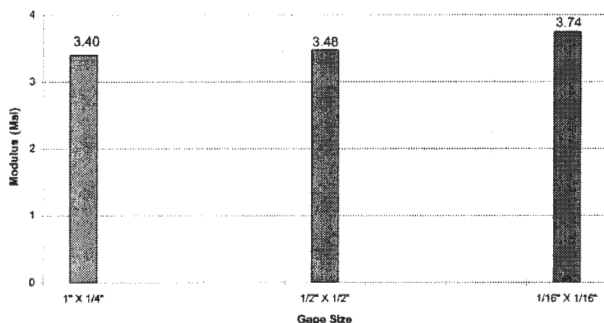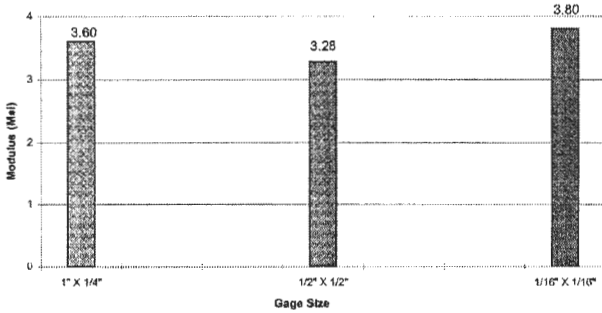

Figure 9: Average Modulus in 0° Fibers Location

Finally, in holding strain gage location relative to a q-cell constant, variation in strain observed at the 90° fiber location was concluded to be significant between the $\frac{1}{16}$" x $\frac{1}{16}$" and $\frac{1}{2}$" x $\frac{1}{2}$" strain gages. That is, the modulus of the $\frac{1}{16}$" x $\frac{1}{16}$" and the $\frac{1}{2}$" x $\frac{1}{2}$" strain gages deviate from each other by an ample amount at this location and should be compared to one another before bonding occurs at the 90° fiber location. The $\frac{1}{16}$" x $\frac{1}{16}$" and the

$^1/_2$" x $^1/_2$" strain gages deviated by more than 20% from the average strain measurement at the 90° fiber location. The smallest amount of variation from the average modulus measurement at the 90° fiber location was observed in the $^1/_2$" x $^1/_2$" strain gage. This strain gage deviated by 6% from the average modulus measurement at this location. The average modulus recorded at this location was 3.54 Msi. The average standard deviation calculated at this location is .214. This can be observed below in Figure 10.

Figure 10: Average Modulus in 90° Fibers Location

When size is held constant, an effect of strain gradients on strain gage size is noticed. In the $^1/_2$" x $^1/_2$" strain gage, the average modulus was calculated as 3.450 Msi. Empirically, the largest amount of deviation in modulus within the $^1/_2$" x $^1/_2$" strain gage was observed at the transverse axis location relative to a q-cell. The smallest amount of deviation in modulus within the $^1/_2$" x $^1/_2$" occurs at the 90° fibers location relative to a q-cell. Statistically, the only significant variation in modulus is noted between strain gages when the 90° fibers location is compared to the transverse axis location. When the 90° fibers location is compared to any of the auxiliary strain gage locations, excluding the transverse axis location, there is no significant variation in modulus. Thus, care must be taken in choosing a strain gage bonding location. The 90° fibers location deviates from the average modulus by 16% of the average modulus, whereas, the transverse axis location deviates from the average modulus by only 10% of the average modulus. This can be observed in Figure 11.
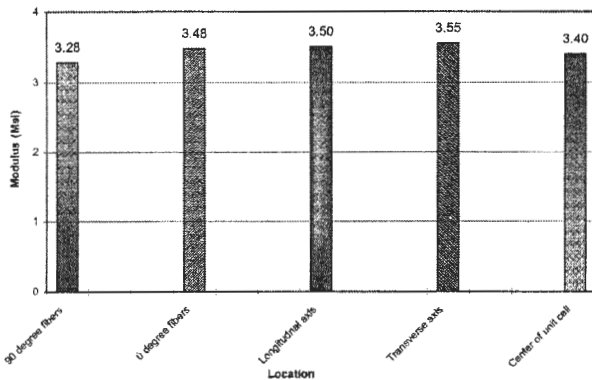
Figure 11: Average Modulus as a Function of Gage Location – Grid Size: $^1/_2$" x $^1/_2$"

In the $^1/_{16}$" x $^1/_{16}$" strain gage, a significant variation in modulus is observed when the center of the unit cell location is compared to the 0° fibers and 90° fibers locations. The largest deviation in modulus from the average modulus is at the center of a unit cell loca-

tion. It deviates from the average modulus by 45% of the average modulus. The least amount of deviation from the average modulus occurred at the 0° fibers location relative to a q-cell. It deviates from the average modulus by 19% of the average modulus. Therefore, when measuring modulus with a $\frac{1}{16}$" x $\frac{1}{16}$" strain gage, the 0° fibers location is preferred when compared to the 90° fibers and center of a unit cell locations. Statistically, there is no significant deviation in modulus when the 0° fibers location is compared to the longitudinal axis or transverse axis locations. This can be observed in Figure 12.
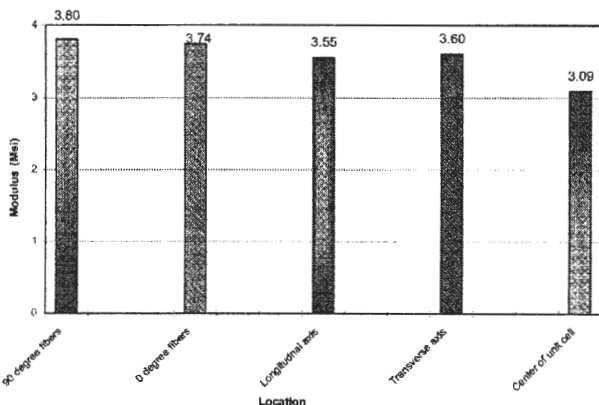
Figure 12: Average Modulus as a Function of Gage Location – Grid Size: $\frac{1}{16}$" x $\frac{1}{16}$"

Finally, the 1" x $\frac{1}{4}$" strain gage shows no significant variations in modulus in a statistical analysis. Empirically, variation in strain is observed at every position, yet deviation in modulus is not significant enough to select one bonding position over another. The largest amount of variation in modulus from the average modulus is located in 90° fibers location. It deviates from the average modulus by 10% of the average modulus. The smallest amount of variation in modulus deviates from the average modulus by .2% of the average modulus. This can be observed below in Figure 13.
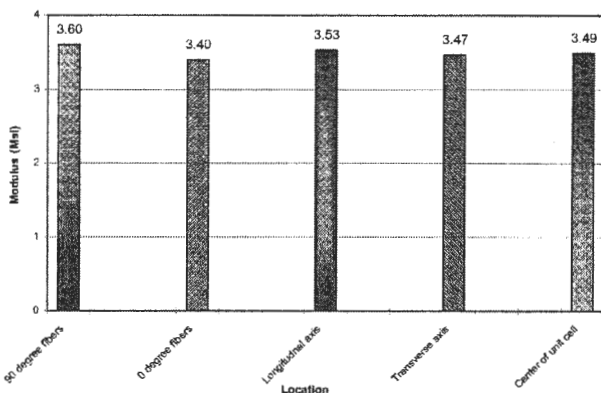
Figure 13: Average Modulus as a Function of Gage Location – Grid Size: $\frac{1}{4}$" x $\frac{1}{4}$"

On average, the least amount of variation in strain is observed in the 1" x $\frac{1}{4}$" strain gages. The most amount of variation is observed in the $\frac{1}{16}$" x $\frac{1}{16}$" strain gage. This type of variation is expected due to the amount of area covered by the 1" x $\frac{1}{4}$" versus the $\frac{1}{16}$" x $\frac{1}{16}$" strain gage. The 1" x $\frac{1}{4}$" strain gage covers four q-cells while the $\frac{1}{16}$" x $\frac{1}{16}$" strain gage only

covers one-sixteenth of a q-cell. Therefore, the larger strain gage will acquire a more representative structural level of strain than the small gage. Overall, the average strain measured in each gage converges to the same value. This activity represents the ability to attain average strain through multiple strain measurements within a strain gradient. That is, the average strain in glass/vinyl-ester can be obtained with any of the previously mentioned strain gages by averaging strain over an abundant amount of data. This activity can be observed in Figure 14. This figure represents 15 gages per strain gage size totaling 75 strain gages overall.
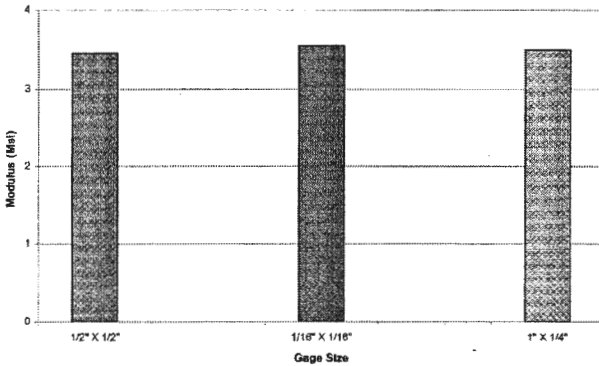
Figure 14: Average Modulus for Gage Size Over All Samples

CONCLUSION

Strain gradients affect strain measurement in a glass/vinyl-ester composite. Local strain measurement is dependent on strain gage size and location. Typically, strain measurement obtained over all specimens and gages approach an average value of strain. Locally, large strain gages give more consistent results than small strain gages. Large strain gages cover more specimen area than small strain gages; thus, large strain gage measurement is validated by an increased amount of strain points located within a strain gradient. Therefore, large strain gages should be used for determining structure-level mechanical properties. The aspect ratio of large gages does not affect measured strain.

Strain measured with small gages indicates that strain gradients exist in a glass/vinyl-ester material. In measuring strain within a $^1/_{16}$" x $^1/_{16}$" strain gage, a large variation in strain is observed relative to position along the specimen. In fact, the amount of variation in strain among all sized gages at all positions relative to a q-cell occur between the 1" x $^1/_4$" and $^1/_{16}$" x $^1/_{16}$" strain gages. The small strain gage ($^1/_{16}$" x $^1/_{16}$") in this experiment measured strain in proximity to a point along the specimen. As a result of strain varying from point to point within a strain gradient, strain is expected to deviate from the average strain.

In addition, mechanical property test programs should account for strain gradient effects in woven composites. That is, when conducting a mechanical property test, one should consider the effects of strain gradients relative to the testing material. In glass/vinyl-ester, strain gradients varied due to position within the sample. For example, a significant variation in strain was observed at the 90° fiber location when this location was compared to the transverse axis location in a $^1/_2$" x $^1/_2$" strain gage.

REFERENCES

[1] Masters, J.E. and P.G. Ifju, "Strain Gage Selection Criteria for Textile Composites", Journal of Composites Technology & Research, Vol. 19, No. 3, pp. 152-167, 1997.

[2] Jones, Robert M., Mechanics of Composite Materials, 2nd Edition. Philadelphia: Taylor & Francis, 1999.

[3] Jenkins, C. H., Manual on Experimental Methods for Mechanical Testing of Composites. Rethel: Society of Experimental Mechanics, 1997.

Belief Fixation: Voluntarism vs. Non-Voluntarism

DeAdera Gardner

Mentor: Dr. Charles Nussbaum, Department of Philosophy

ABSTRACT

There are many different views that accompany the question, "Is belief voluntary?". One of the problems with this question is that many philosophers do not make the necessary distinctions concerning voluntarism. Close scrutiny has shown that there are two types of voluntarism. The two types of voluntarism are epistemic responsibility and voluntarism as action. There are also distinctions to be made between direct and indirect action and dispositional and occurrent beliefs. Upon reflecting on this information, the research shows that voluntarism as a mere act is simply an impossibility. Contrary to this, voluntarism as epistemic responsibility is very plausible and because of this, I conclude we are epistemically responsible for our beliefs.

INTRODUCTION

There are many issues in life that go unexamined. Some of these things may be viewed as trivial and others as significant. One such issue is the question, "Is belief voluntary?". Before this question can be answered intelligently, there are a number of distinctions that must be made. Most important is the distinction between two types of voluntarism. The first holds the adoption of belief to be an action. This attributes to a person the ability to believe a certain proposition at will, in the same way a person may have the ability to blush at will. The other type of voluntarism is identified as cognitive or epistemic responsibility: the ability to decide what to believe based on evidence and to be held responsible for one's beliefs. Next, there is a distinction to be made between the adoption of a belief as a direct or indirect action. A direct action is something that can be immediately induced, whereas an indirect action has intermediate steps that must be taken in order to produce the desired action. Lastly, there is an important distinction between occurrent and dispositional beliefs.

I. Belief Voluntarism as Action

With these distinctions in mind, I may begin to explore the question of whether or not belief is voluntary by attempting to understand what is involved when viewing voluntarism as the adoption of a belief at will. There are many philosophers who hold that a person can will herself to believe a particular proposition if the belief state is achieved through some *indirect* means. For example, if a person goes to see a hypnotist in order to bring about a certain belief, the belief was achieved through an indirect action. Louis Pojman (1985) has presented two arguments against belief volition as *direct* action. The first is the Phenomeno-logical Argument, which goes as follows:

1. Acquiring a belief is a happening in which the world forces itself upon a subject.
2. Happenings in which the world forces itself upon a subject are not things the subject does (i.e., are not basic acts) or chooses.
3. Therefore, acquiring a belief is not something a subject does (i.e., is not a basic act) or chooses (1985, 40).

The main point Pojman is trying to convey in this argument is that acquiring a belief is something that just seems to happen and does not feel like a matter of choice. One can choose to entertain a proposition at will at any given time, but one cannot choose to believe a proposition at will.

Pojman's second argument, the Logic of Belief Argument, deals specifically with the conceptual incoherence of volitionalism. The argument is as follows:

1. If A believes that p, A believes that p is true (analysis of the concept of belief).

2. In standard cases of belief, the truth of p is wholly dependent on the state of affairs s, which either corresponds to p(and thus makes p true) or does not (and thus makes p false).[1]

3. In standard cases of belief, whether or not the appropriate state of affairs s that corresponds to p obtains is a matter that is independent of A's actions and volitions, but there is some truth connection between s and p (from 2).

4. In standard cases of belief, A subconsciously or consciously believes or pre-supposes premise 3.

5. Rational believing is defined as believing according to the evidence and fully rational believing as believing simply because of the evidence.

6. Therefore, in standard cases of belief, A cannot fully rationally both believe that p and that his belief is presently caused by his willing to believe that p. Rather, if rational, A must believe that what makes his belief true is a state of affairs s which obtains independently of his will (1985, 48-9).

In short, Pojman argues that one cannot possibly believe a certain proposition and at the same time truly believe that it is caused by his willing to believe the proposition. Pojman says, "Volitional believing is not simply irrational believing, but it is incoherently irrational, for it offers an account of believing that confuses the nature of believing…Hence we can say that doxastic incoherence is a species of epistemic irrationality" (Pojman 1985, 49).

There are other philosophers who address the issue of belief voluntarism as an act. Barbara Winters (1979) argues for a similar, but weaker, conclusion than Pojman's.

It is impossible for me to believe of a particular belief b that b is a present belief of mine and is sustained at will…[This] does not entail that I cannot believe I have acquired beliefs at will, or that I cannot regard b, which I acquired at will, as a present belief of mine, or that I must stop believing b if I come to think b is sustained at will (Winters 1979, 256).

What is entailed for Winters is that an individual cannot believe that she believes b and believes that b is sustained merely by an act of will. For Winters the impossibility of belief voluntarism involves second-order beliefs. A belief about a belief is a second-order belief. Winters introduces second-order beliefs terminology in order to allow for uncon-scious beliefs that were acquired by will. Any belief about which a person has a second-order belief cannot, however, be unconsciously held, since she believes that she holds that very belief. Although Pojman criticizes Winters' conclusions as too weak, to decide this issue would require determining whether unconscious beliefs are possible, which is beyond the scope of this paper.

II. Belief Voluntarism as Epistemic Responsibility

If belief voluntarism is not viewed as action, it is viewed as epistemic responsibility.

---

[1] The actual text contains the misprint: "In standard cases of belief, the truth of p is wholly dependent on the state of affairs s, which either corresponds to p (and thus makes p false)".

Réné Descartes (1986) is one philosopher who takes the latter view of belief voluntarism. He says that individuals are epistemically responsible for their beliefs because he attributes to humans the ability to suspend judgment in situations where there is not enough evidence to decide one way or the other. Descartes holds that if a proposition is entirely clear and distinct, belief legitimately arises and one is bound to accept the proposition. He says the will is functioning properly when in accord with clarity of intellectual insight. His view differs slightly from that of others because to be free for Descartes means a believer is compelled to follow what rationality tells her. By this line of reasoning, one is free only when judging rationally because that is what her true self is saying. For Descartes (paradoxically) to be bound by reason is to be free. To judge capriciously or groundlessly, however, is to be passively determined or unfree.

Descartes' position is an epistemologically internalist one. Epistemological internalism is the view that a person is epistemically justified only if she actually gives or is able to give reasons that warrant belief in the proposition. On the other hand, externalism "requires only that the belief in the proposition have the right sort of causal ancestry – for example, that it be a product of a mechanism that generally results in true beliefs" (Clarke 1986, 39). Both internalism and externalism have their pros and cons. Internalism comports well with epistemic responsibility, but internalism has to deal with the Gettier problems (Gettier, 1963/1995), which question the traditional definition of knowledge as justified true belief. Imagine someone has a friend who drives a Ford. A neighbor tells her, "I saw your friend yesterday. That is a nice vehicle she has. What kind is it?" She says it is a Ford. Unbeknownst to her, her friend just bought a new car the day before, but luckily it is also a Ford. In this situation the individual is in possession of justified true belief, but she still does not *know* her friend has a Ford. Her statement, though justified and true, is based on a lucky coincidence. What is required is a reliable connection, whether causal or otherwise, between an individual and the world. The externalist view is that this connection does not have to be known to be reliable in order to be reliable. But this seems to rob the knower of epistemic responsibility, since the fact that the connection is reliable need not be available to her or him. Internalism on the other hand, makes sense of epistemic responsibility, but it is less effective in dealing with the Gettier problems.

Hilary Kornblith is an externalist who resists the conclusion that externalism is incompatible with epistemic responsibility. He believes that, while not all of a person's cognitive states are subject to control, she has the ability to influence her own epistemic habits. Kornblith relies on what he calls the Principle of Epistemic Responsibility. The principle says that a person is responsible for the beliefs she has, and, if that person has an unjustified belief, that person is epistemically culpable. This principle says that the person is responsible for cognitive states accessible to her, as well as for states inaccessible to her. But Kornblith asks, "How can one reasonably hold a person responsible for the transitions that person makes among mental states which he doesn't know about?" (1982, 243). Kornblith grants that, even if the principle is true and we are responsible for the beliefs we have, this does not require an internalist view of knowledge.

Here is where Kornblith's externalist epistemological theory comes into view. He points out that one does not have to know the inner workings of a machine in order to judge whether it is functioning properly. The same is true of belief mechanisms. It is impossible for each individual knower to develop a complete theory of psychological processes in addition to a complete theory of what makes such processes justification-conferring. He

says, "All we need to do is monitor the output of our psychological processes. This amounts to nothing more than being on the lookout for false beliefs" (Kornblith 1982, 245). This process does not require that anything be known about the manner in which the beliefs were produced. Individuals can change their epistemic habits without knowing how a belief mechanism works.

III.   Content of Belief

To determine whether or not belief is voluntary, one must also consider the content of belief or what the belief is about. Belief content can concern both matters of fact and values. William James says (1917) that as believers we generally are faced with three types of options or choices between hypotheses. An option can be living or dead, forced or unforced, and momentous or trivial. A living option is a hypothesis that appeals as a real possibility to whom it is proposed. A dead option is one that makes no connection with one's nature. An example of a dead option for most people today would be the hypothesis that witches exist. A forced option means one must choose or reject a particular proposition. An example of a forced option is the belief in the existence of God because suspending judgment is like rejecting the hypothesis. James says if this is one's option, and there is no evidence one way or the other, one may will to believe. James says a genuine option is one that is living, forced, and momentous. Such options tend to concern values.

According to James there are also times when willing to believe is what is necessary to make something true. The example he gives is of someone wondering about the status of the relationship between herself and another person. She wonders if the other person likes her. When one is willing to believe that the other person likes her, she acts in a friendly and open way, and it could be that very action and will to believe that cause the other person to like her. Similarly, to act withdrawn and cold towards the other person could cause the person to dislike her. For James, willing to believe is sometimes possible and legitimate. He does not claim, however, that such willing to believe is a direct action.

IV.   Beliefs, Assertions, and B-states

Another issue that must be addressed before determining for or against voluntarism concerns beliefs, related cognitive states, and speech acts. First, we distinguish what Bernard Williams (1973) calls a belief and a B-state. Williams attributes three characteristics (among others) to belief. He says beliefs aim at truth, can be based on evidence, and need not be asserted. Like Pojman, Williams concludes that one cannot simply will to believe a certain proposition because beliefs aim at truth. "If in full consciousness I could will to acquire a 'belief' irrespective of its truth, it is unclear that before the event I could seriously think of it as a belief, i.e., as something purporting to represent reality" (Williams 1973, 148). Williams introduces the notion of a B-state, a state of a cognitive system or machine that possesses the first and second characteristics of a belief, namely that beliefs aim at truth and can be based on evidence. In order to have a belief, Williams thinks, one must be able to assert one thing and think another (or have the ability to deceive). He says that this cognitive machine will not manifest beliefs, or simply put, it would not be considered to have belief states. It will only have B-states. According to Williams, the reason a B-state falls short of belief is that the machine is incapable of making insincere assertions. Knowledge for Williams is the more primitive state. He says one must be able to state what she does not believe in order to have a belief.

V.   Degrees of Confidence, Assent, and Acceptance

Ronald De Sousa (1971) introduces an important distinction between belief as degree of confidence and belief as assent. Belief for De Sousa is defined as what you are willing to assent to or take as true. The logic behind this is that De Sousa says truth is an all-or-nothing affair, where probability has different degrees. Because of this, what a person is willing to take as true will not only be a function of the subjective probability that one attaches to a proposition. Subjective probabilities attach to dispositional beliefs because dispositional beliefs are held with varying degrees of strength. Occurrent beliefs are the beliefs that you are presently acting on. Dispositional beliefs are dormant; they are beliefs that are not being acted on at the moment, as well as beliefs that have never been expressed. An assent for De Sousa is a kind of occurrent belief. De Sousa says a person starts out with a dispositional belief, which has degrees of probabilities. Next, one may choose to assent to the proposition. Once a person assents, the dispositional belief becomes an occurrent one, one that she is prepared to defend.

The main point is that one can believe a proposition without assenting, but if one assents to a proposition, she must be prepared to defend it. Dispositions figure in the explanation of Bayesian Decision Theory in which the most rational action combines high probability with a high utility of outcome. Acceptance, on the other hand, has to do with truth and argument. Assent is only one kind of occurrent belief because occurrent beliefs can play a role in rational decision making as well as in assent. It is clear that dispositional beliefs are not voluntary in any sense. The reasoning behind this is that dispositions are habits, and habits are proclivities, not actions. Some dispositional beliefs may be unconscious and may not be expressed at all. One can surely say a person cannot voluntarily will her unconscious beliefs. Can one decide to assent at will? The answer is "no" because individual assent aims at truth and is based on reasons or observations.

Following De Sousa, Mark Kaplan (1981) introduces a related distinction between confidence and acceptance and, like De Sousa, criticizes the belief as confidence view. Kaplan's first claim is that on the belief as confidence view, a rational person does not have to be certain that a proposition is true before she can accept it. She just has to be sufficiently confident. Kaplan asks the reader to imagine that "1" represents the maximum degree of confidence, which means one is certain that a particular proposition is true. The halfway mark would be at ".5" and "0" would represent total lack of confidence. Lastly, P represents the proposition. With this in mind Kaplan derives his first claim as follows (1981, 130):

1.   There is a number n, where n is greater than or equal to .5 and less than one, such that, if X is rational, then X will accept P if and only if X has a degree of confidence greater than n that P.

This yields necessary and sufficient conditions for acceptance:

For a rational person a degree of confidence (greater than .5) is **sufficient** to warrant acceptance.

1a.   There is a number n, where n is greater than or equal to .5 and less than one, such that if X is rational, then X will accept P if X has a degree of confidence greater than n that P.

For a rational person a high degree of confidence is **necessary** for acceptance to be warranted.

1b. There is a number n, where n is greater than or equal to .5 and less than one, such that if X is rational, then X will accept P only if X has a degree of confidence greater than n that P.

Kaplan's claim is that if a rational person accepts a proposition then that person is also subject to a set of consistency constraints. These are as follows (1981, 131):

2. If X is rational then

a. X will accept the conjunction of any propositions she accepts;

b. X will accept all the consequences of every proposition she accepts; and

c. X will not accept any contradiction.

Now, claim 1a says a high degree of confidence is **sufficient** to warrant acceptance. But there is a case in which this generates a puzzle known as the lottery paradox. Imagine a fair lottery of 100 tickets with one winning ticket. The rational degree of confidence in the proposition that any given ticket will lose is .99, but we do not believe each and every ticket will lose for we know one will win. Because of this, degree of confidence is not sufficient for acceptance.

A second puzzle concerns the claim (1b) that for a rational person, a high degree of confidence is **necessary** for acceptance to be warranted. The example given is known as the historian paradox. Kaplan asks the reader to suppose that X is the historian. After a lengthy time spent on research, the historian has written a rather large book. "As she reads over her magnum opus she reaffirms, for each sentence in the book, her acceptance of the proposition which that sentence expresses in English" (Kaplan 1981, 132). After she has finished reading the book, she realizes that due to human fallibility it is very likely that her work is erroneous in some detail. Because of this, she has a very low degree of confidence (below .5) that the conjunction of all the propositions in the book is true. Because she is rational and she accepts each proposition individually, she should be willing to accept the conjunction of propositions (from the rationality constraints). However, she does not accept the conjunction as true because her degree of confidence that the entire book contains no errors is below .5, and this is where the contradiction lies.

Kaplan mentions three ways to resolve the puzzles. He says one can say that claim 1 is too permissive. To say this is to say that a rational person should accept a proposition only if she is *certain* it is true. The application of this to the lottery case allows the person to profess that she does not believe any ticket will lose. This, however, is not a good option because, if science never accepted any proposition that is less than certain, our knowledge would be very limited. In the case of the historian, adopting this tactic will keep her from accepting much of what she wrote, since she is absolutely certain of little or none of it. Our second option is to deny one or more of the three rationality constraints. But this would undermine the force of all reductio arguments. This is important because a reductio argument derives a contradictory conclusion from a conjunction of propositions, which shows at least one proposition must be false. Reductio arguments are very important for evaluating claims. Kaplan says our final option is to admit that 1a and 1b are incorrect. No degree of confidence is necessary or sufficient for acceptance. In the lottery, a degree of confidence greater than .5 is not sufficient for acceptance. In dealing with the historian paradox, a degree of confidence greater than .5 is not necessary to warrant acceptance.

Kaplan believes most philosophers do advocate the confidence-threshold view of

acceptance. This view holds that to accept a proposition is to have just a high degree of confidence above a threshold short of certainty. Kaplan argues that acceptance is not a state of confidence at all. By this Kaplan is suggesting that, when one is talking about acceptance, she is not talking about confidence or rational decisions but about what she is prepared to defend as true. He is talking of argument, not rational decision.

CONCLUSION

I have found that belief voluntarism involves a set of distinctions which I will review. First, there exist the distinctions between the different types of voluntarism: action vs. epistemic responsibility. Second, it is necessary to keep in mind the distinction between direct and indirect actions. Third, it is important to consider the implications of the Phenomenological and Logic of Belief Arguments. Fourth, one must not forget that the content of a belief can concern both matters of fact and value and that this can bear on the issue of belief voluntarism. Fifth, it is essential to take into account the distinctions between dispositional and occurrent belief, confidence vs. assent or acceptance, and assertion.

Is belief voluntary? I conclude that dispositional beliefs are like habits and therefore not voluntary. Although one cannot decide to have a certain disposition, one can train the mind by repeatedly taking steps to induce the desired result. Perhaps it is something like this that James had in mind. Because assertion is an action (a speech act) one can will to assert. But unlike assertion, one cannot simply will to assent because assent is based on reasons or on observations. The answer, then, to the question of whether belief fixation is voluntary or involuntary is that it depends: it depends on which type of belief voluntarism you are speaking of. Belief voluntarism as direct action is not possible, but belief voluntarism as indirect action is possible: we are able to modify our beliefs by indirect means. In addition, belief voluntarism as epistemic responsibility is, I think, a very plausible notion.

What, if any, practical implications does this have? The fact that individuals are epistemically responsible for their beliefs says a great deal. It is obvious that if people choose to believe something false or incorrect, then they are being epistemically irresponsible. In light of this, I can also say that individuals can and should develop the habit of recognizing and accepting true beliefs. The foregoing considerations have shown that belief voluntarism as direct action is false, but as indirect action and as epistemic responsibility, it is true.

REFERENCES

Bennett, J. (1990). "Why is Belief Involuntary?" Analysis 50: 87-107.

BonJour, L. (1980). "Externalist Theories of Empirical Knowledge." Midwest Studies in Philosophy 5: 53-57.

Clarke, M. (1986)."Doxastic Voluntarism and Forced Belief." Philosophical Studies 50: 39-51.

Descartes, R. (1986). Discourse on Method and Meditations on First Philosophy. Cambridge, New York: Cambridge University Press.

De Sousa, R. (1971). "How to Give a Piece of Your Mind: or the Logic of Belief and Assent." The Review of Metaphysics 25: 52-79

Gettier, E. (1963/1995). "Is Knowledge True Justified Belief?" In Empirical Knowledge: Readings in Contemporary Epistemology, ed. Moser, 237-260.

Harman, G. (1982). "Positive vs. Negative Undermining in Belief Revision." In Naturalizing Epistemology, ed. Kornblith, 317-336.

Heil, J. (1983). "Doxastic Agency." Philosophical Studies 43: 355-364.

James, W. (1917). "The Will to Believe" In Other Essays in Popular Philosophy. New York: Longmans, Green, and Co.

Kaplan, M. (1981). "Rational Acceptance." Philosophical Studies 40: 128-145.

Kornblith, H. (1982). "The Psychological Turn." Australasian Journal of Philosophy 60: 238-253.

Kornblith, H. (1995). Naturalizing Epistemology (Second Edition). Cambridge: MIT Press.

Moser, P. (1996). Empirical Knowledge: Readings in Contemporary Epistemology (Second Edition). Lanham, Maryland: Rowman and Littlefield.

Moser, P., Mulder, D. and Trout. J.D. (1998). The Theory of Knowledge: A Thematic Introduction. Oxford: Oxford University Press.

Peirce, C. (1978). "The Fixation of Belief." in The Philosophy of Peirce: Selected Writings, ed. Justus Buchler. New York: AMS Press.

Pojman, L. (1985). "Believing and Willing." Canadian Journal of Philosophy 15: 37-56.

Pollock, J. and Cruz, J. (1999). Contemporary Theories of Knowledge (Second Edition). Lanham, Maryland: Rowman and Littlefield.

Quine, W. and Ullian, J. (1978). The Web of Belief (Second Edition). New York: Random House.

Williams, B. (1973). "Deciding to Believe." In Problems of the Self: Philosophical Papers. New York: Cambridge University Press.

Winters, B. (1979). "Believing at Will." Journal of Philosophy 76: 243-256.

Digital Superimposition Techniques in Skull/Face Comparison, Fingerprint, and Handwriting Analyses: Applications for Medicolegal Professionals

Tamara Dawn Robertson

Mentor: Dr. Dana Austin, Sociology and Anthropology

ABSTRACT

This paper surveyed the basis of forensic media superimposition, which is the craniofacial identification technique. The application of the methodology that has its beginnings in craniofacial superimposition was explored for handwriting analysis and, in more detail, for fingerprint comparison. Literature pertaining to skull-to-face superimposition, questioned document analysis and superimposition, and fingerprint analysis was reviewed; current methodologies were outlined and examples given. A preliminary experiment in fingerprint superimposition was designed and implemented. Bifurcations and ridge endings were evaluated in superimposed comparisons of ten sets of thumbprints, and when evaluated by a t-test for means, results showed that bifurcations yielded a confidence of 99.999 percent in matching, and ridge endings yielded 99.995 percent confidence using three of each minutia. When the minutiae were combined for a total of six, the prints were matched with nearly 100 percent confidence.

> *In the unbiased search for the truth the law has no favorites by presumption. Silent circumstances, without the power to change their attitude, or to make explanations, or to commit perjury, may speak as truthfully in court as animated witnesses. ... If the truth is found in oral testimony, it must determine the issue, but it is equally potent if found in circumstances.* (Justice Rose, quoted in Osborn 1929, p. x)

INTRODUCTION AND SETTING OF PROBLEM

Since the advent of organized systems of law humankind has sought more efficient ways of achieving justice by means of identification of victims and perpetrators of crimes. The successful identification of a victim or missing person assumes paramount importance, not only to professionals in the fields of law and medicine, but also to loved ones in need of information and closure (Burns 1999). Moreover, to positively identify the perpetrator of a crime is to lay a foundation for a true system of justice, rather than simply settling for a system of compromise. The use of media superimposition as a viable tool in furthering these aims has greatly advanced under the guidance of both biological and forensic anthropology (İşcan 1993a). However, few methodologies that approach standardization have been attempted for digitized photographic superimposition (İşcan 1993b), nor has this technology been made readily accessible to law enforcement and forensic medical teams that must operate under the constraints of restricted budgets. The implementation of consistent procedures will benefit the entire field of identification services by making superimposition techniques available to medicolegal professionals who may have a wide range of experience with both visual media and computer equipment. In many cases, photographs of missing persons may be available even if no medical or dental records exist, and the angles of the photographs taken before death (antemortem) can be carefully reconstructed for comparision with photographs or video footage taken after death (postmortem) (İşcan 1993a). Superimposition techniques can also be applied to other aspects of forensic comparison, such as handwriting and fingerprint analyses, although the criteria differ somewhat.

Superimposition of visual media as a technique for craniofacial and questioned document identification has come to be accepted as legal evidence (Nickerson et al. 1991; Osborn

1929). Identification standards, for example, have been developed in China for several large minority nationalities (Lan 1995). In this country, however, we have not begun to thoroughly explore the possibility of using digitized technology as a widespread method for the identification of victims and criminals. Drs. Dana Austin and William Maples conducted a comprehensive study proving that video based craniofacial superimposition is scientifically significant as a method for identification in the United States (Austin-Smith and Maples 1994). Using photographs taken from frontal and lateral views in a video mixing system has been shown to be over 99 percent effective. Medicolegal professionals have an obligation to employ every available tool to procure the highest possible standards ensuring that positive identifications (IDs) can be made. Accordingly, this research will explore the advancement of standards for digitized media superimposition as it applies to skull/photo, questioned document, and fingerprint analyses.

This paper surveys the effectiveness of image superimposition as a technique to aid forensic anthropologists in the task of postmortem identification. This paper also proposes that this tool may also be applied to many aspects of forensic and criminal investigations. Additional applications include handwriting and fingerprint superimposed comparisons and analysis. The latter was the subject of a preliminary study to ascertain its effectiveness, as a description of this technique in regard to forensic laboratory analysis has never been published. Two problems have been encountered in the course of this research. First, a strategic and comprehensive methodology must be synthesized and amended from models of skull-to-face and questioned document superimposition, using the manipulation of digital imaging, for use as a guideline in implementing the techniques for fingerprint superimposition. Second, sample cases must be statistically analyzed and interpreted for fingerprint comparisons in order to evaluate the effectiveness of the new procedures. Analysis of superimposition will rely on previous research to validate the claim that digitized craniofacial superimposition and photographic questioned document superimposition stand as scientifically reliable and valid techniques (Lan 1995, Osborn 1929, Nickell 1996).

Research was conducted under the assumption that media superimposition is an effective, if not universally accepted, technique for identification purposes. Thus, the null hypothesis that was tested in this research states that approximately 50 percent of the thumbprint samples tested should accurately yield a consistent fit, indicating that they are from the same individual. This value is similar to the chance outcome of approximately 10,000 coin tosses or more (Sims personal communication), which supports the assumption that the null hypothesis is invalid. Two research hypotheses have also been constructed prior to conducting this study. The first states that the use of anthropological and criminological criteria will enhance the effectiveness of the technique. This hypothesis ensures a valid and reliable research design by adopting the standards regularly used in these fields. The second states that at least 99 percent of cases should accurately yield a consistent fit. This hypothesis ensures the scientific integrity of the preliminary findings while allowing some margin to avoid committing a Type I error that would falsely exclude an individual. In addition, this study published sensitive case information only with the expressed permission of the parties involved and/or their next of kin. The research design was also limited to image superimposition techniques involving photographic media adapted for use with digital editing software.

Specific definitions relating to the methods employed in comparative identification must be operationalized. The two types of personal identification (ID) referred to in this

research are tentative and positive (Austin personal communication). Tentative refers to an unconfirmed ID of an individual based on investigative leads. Positive describes a verified ID of an individual based on scientific methods and/or viewing by immediate family members. For craniofacial superimpositions, several measurements may be used to determine how consistently the skull image fits with the facial photograph(s) (Lan 1995, Işcan 1993b). Craniometric indicators include biorbital width, bigonial width, interorbital width, intraorbital width, bizygomatic width, nasal width, intraorbital height, nasal height, upper facial height, and total facial height (Burns 1999); some of the points on the face used for these measurements are illustrated in Figure 4. Questioned document analysis is defined as the scientific analysis of handwriting and the comparison of documents, including signatures, to determine authenticity and is distinct from its predecessor, graphology (Hilton 1982). Dactylography refers to the analysis of fingerprints and is distinct from dactyloscopy, the classification of fingerprints (Foote 1974).

REVIEW OF THE LITERATURE

Although the technique of superimposition has incorrectly been referred to as "facial reproduction" (Caldwell 1986, p. 229), it is much more likely to have evolved from issues arising from comparative identification (Krogman and Işcan 1986). Published accounts of comparisons of skulls with busts, portraits, and death masks predate the beginnings of superimposition in the nineteenth century (Grüner 1993). Glaister and Brash reported the earliest use of superimposition by a European researcher in 1867 (Aulsebrook et al. 1995), and the method has slowly matured and spread in use throughout the twentieth century. Grüner (1993) reports early superimpositions, by Welcker in 1883, of portrait sketches of famous figures, including the philosopher Immanuel Kant. He also relates the recasting of a superimposed composite bust of the famous classical composer, Johann Sebastian Bach, by His in 1895. The advent of photographic technology in the nineteenth century resulted in more sophisticated techniques that enabled forensic scientists to be more precise (Foote 1974). Glaister (1973) recounts the famous Ruxton case of 1935, as well as the Dobkins case of 1942, that used craniofacial superimposition in the process of identifying the victims. In the former case, Dr. Ruxton murdered his wife and her maid, scattering their remains over a wide area to confuse the authorities. However, most of the remains of both women were recovered, including the skulls. Professor J. C. Brash then superimposed a negative of one of the skulls recovered onto a portrait of Mrs. Isabella Ruxton, shown in Figure 1, and achieved a consistent fit that confirmed other available evidence.

Figue 1: Superimposition of Negative of Skull onto Portrait of Mrs. Isabella Ruxton (from Glaister 1973)

The Dobkins case was similar to the Ruxton case in that Mr. Dobkins murdered his wife and attempted to obfuscate the evidence by dismembering her body. Photographic superimposition was employed in the identification of the remains, in addition to the confirmation by examination of an antemortem uterine tumor and dental fillings. Since these early cases, three stages of superimposition have been delineated (Aulsebrook et al. 1995, p. 102):

- The first made use of viewing boxes, slide projectors, and overhead projectors.
- The second phase introduced video technology and its capacity for a variety of electronic superimpositions and comparisons.
- The third incorporated the analytical potential of computer graphics.

The final stage is the focus of the following discussion.

Perhaps the best study to exemplify the methodologies used in digitized photographic superimposition is that of Yuwen Lan, in which living subjects were radiographed and photographed. Lan superimposed the two images and determined fifty-two indices from ten "determining lines," landmarks, soft tissue indices, and "index numbers" (1995, p.137). Lan used his findings to develop identification standards for each of fifteen test nationalities in China and incorporate the information into a software package he designed with Dongsheng Cai. He then performed a statistical test to determine the validity of the results, which proved beyond a doubt that the indices created were highly significant for all test nationalities. He concluded from this study that each nationality could be identified from the distinct landmarks, soft tissue indices, and index numbers that were defined and that the indices were unique to their respective nationalities. Yoshino et al. also used a computer aided photoanthropometric video superimposition approach to determine consistency of fit for "craniofacial superimposition images" (1995, p. 129). Their data differs slightly from that of Lan (1995), in that they combined morphological data with tissue depths and anthropometric data that they referred to as "positional relationships" (p. 132). Using what appears to be a form of statistical analysis called a sign test, they evaluated the consistency of the superimpositions using a total of eighteen criteria. The authors (Yoshino et al. 1995) of this study

concluded that superimpositions in which at least thirteen criteria were "anatomically satisfied" resulted in positive IDs (p. 133). With the help of other researchers, they later proposed an actual video superimposition system incorporating the use of computer digitization (1997). Bajnóczky and Királyfalvi (1995, p. 160) used a computer assisted video system to advance craniofacial superimposition. Their approach measured and marked points and used data matrices to determine the "tolerance limit" for deviations to prevent a Type I, or false exclusion error. The method they employed has both advantages and disadvantages, but they stipulated that this type of system should be used in conjunction with video superimposition "as an independent check," due to the possibility of chance error and limitations arising from comparison of manual markers with pixels (p. 160).

Questioned document analysis was employed in ancient Rome as early as the first century and was discussed by great scholars, such as the rhetorician Quintilian (Osborn 1929), although graphology was also a standard practice from which questioned document analysis may have developed (Nickell 1996). In fact, a portion of an ancient Roman text by Quintilian combined the question of document authenticity with the assertions that the handwriting made about the creator of any given document. Nickell (1996) states that a well-known Chinese philosopher in the eleventh century endorsed the practice of graphology. From this period until the seventeenth century "little is known about the history of graphology" (Hargreaves and Wilson 1991). The first published treatise appears to be a work by Dr. Camillo Baldi in 1622, and the first work of "modern graphology" was published in France in 1872. Since the late nineteenth century, questioned document analysis has become an established "science and art" (Osborn 1929, p. vii). Although questioned document analysis used for evidentiary purposes does not incorporate the interpretation of an individual's personality, which is the definition of graphology (Hilton 1982), both fields agree that people develop certain patterns in their handwriting that generally deviate slightly about the *model patterns* (Osborn 1929, Hilton 1982, Hargreaves and Wilson 1991, Slyter 1995). Osborn (1929) noted that these deviations varied with age and amount of time spent writing. He also noted that abnormal circumstances, like coercion or illness, could affect a handwriting sample. Superimposition has been employed for the purpose of detecting forgeries because of the convenience of comparison (Osborn 1929, Hilton 1982, Nickell 1996). Figure 2 shows a photographic superimposition of a document determined to be a forgery onto a genuine document. The near perfection of the letter forms in this signature should be noted.
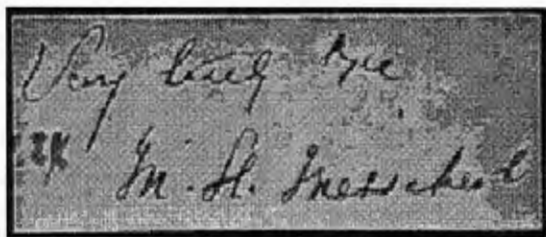
Figure 2: Superimposition of Questioned and Authentic Signatures (from Osborn 1929)

The use of fingerprints dates back even further, to a period more than 2,200 years ago (Chapel 1941). An early Chinese practice included the fingerprint of a person as part of the signature in a seal. In addition, Chinese contracts in the seventh and eighth centuries included the individual's nail mark or fingerprint (Galton 1892, Chapel 1941). Fingerprints

were historically employed on official government documents as early as the fourteenth century in Persia (German 2000). In the seventeenth century, scientists became interested in ridge patterns on both the fingers and palms, resulting in at least three scientific treatises (Foote 1974). J. C. A. Mayer was the first researcher to strongly assert the individuality of fingerprints, indicating in 1788 that despite the "certain likeness" of human fingerprints, similarities among individuals were not strong enough to conclude that any two were the same (Foote 1974, p. 2). However, not until the late nineteenth century were these patterns studied for the scientific purpose of identification. An American microscopist, Thomas Taylor, gave a lecture in 1877 suggesting that fingerprints and palm prints might be valuable for criminal identification (Foote 1974). Shortly thereafter, Scottish surgeon Dr. Henry Faulds made the first fingerprint ID from a latent print taken from a bottle (German 2000).

Due to the dedicated efforts of Dr. Faulds and others, the science of dactylography, or fingerprint analysis, was born. After failing to enlist the aid of Sir Charles Darwin, the doctor was referred to Darwin's cousin, Sir Francis Galton, who published the first major treatise on fingerprint patterns and their uses in 1892 (German 2000). Galton included in his work the uses of fingerprints for personal and criminal identification (Galton 1892). He also identified the three basic pattern types still used today: arches, loops, and whorls. Arches are a pattern in which "the ridges run from one side to the other... without making any backward turn or twist," while loops have "a single backward turn, but no twist." It should be noted that the formation of a double loop, or twinned loop, also occurs (*The Science of Fingerprinting* 1985). Galton (1892) further pointed out that whorls have "a turn through at least one complete circle." Although the terminology in the definitions has changed greatly over time, the basic definition is still the same (*The Science of Fingerprinting* 1985). Since that time, dactylography has become a nearly universal practice that has evolved into classification systems; however, no superimposition techniques have been found in the literature that could benefit medical and law enforcement personnel (Galton 1892, Chapel 1941, Foote 1974, Olsen 1978, *The Science of Fingerprinting* 1985). Figure 3 exemplifies the current methods used to compare prints. The corresponding numbers in the two charts (latent and inked fingerprints) refer to points of comparison that are called minutiae. The minutiae are defined as "minute peculiarities," or characteristics, in ridges (Galton 1892).
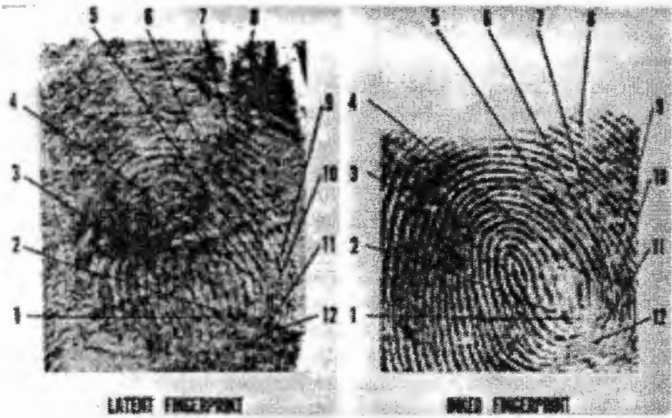
Figure 3. Fingerprint Comparison Chart (from Olsen 1978)

RESEARCH DESIGN

The materials, software, and media employed for the purposes of this study must be delineated. The principal researcher scanned in the craniofacial example from Glaister (1973) and fingerprints using the Astra 2000P reflective flatbed scanner and a Pentium II desktop computer with 128 megabytes of memory. The handwriting samples were prepared for an Internet site by another researcher (Brugnatelli 1999) and downloaded from the Internet. The software used in conjunction with the scanner was Presto! PageManager by New Soft, and all photographic media were edited and manipulated using Adobe PhotoShop version 5.5. The principal researcher analyzed craniofacial superimpositions in JPEG format as recommended by the software. The set of handwriting samples were downloaded from the Internet in GIF format, and they were also analyzed in that format. The two-dimensional scanner was used to digitize the fingerprints for the statistical analysis at 9600 dpi (5 percent size), and they were analyzed in PSD (Adobe Photoshop) format.

For craniofacial superimpositions, methodologies involving expensive video systems can be simplified using either a camera and scanner or a digital camera, which include the same type of viewfinder that is found on most hand-held video cameras. The angle of rotation, called "deflection," and the angle of vertical tilt, called "pitch," can each be indexed using three anthropometric points for each angle (right and left ectoconchions to glabella and glabella to subnasale to gnathion, respectively) on the photograph (Lan and Cai 1993, p. 119). Using these indices a forensic scientist can precisely determine the position for the skull from a photograph using positioning devices no more complicated than a cork flask ring and protractor. Once the correct positioning has been achieved for the skull, the final photographs of the skull and face can be digitized for superimposition, using editing software that can layer photographs. A variety of morphological features and photo-anthropometric points can be used for the superimposed comparison to ensure the consistency of fit. Işcan (1993b) enumerates several features and points that can be used in the course of this comparison. Austin and Maples (1994) also describe morphological features that parallel Işcan's (1993b) data. Examples of craniometric points that are commonly used in craniofacial superimposition, including those used by Lan and Cai (1993) to index deflection and pitch are shown in Figure 4 below.
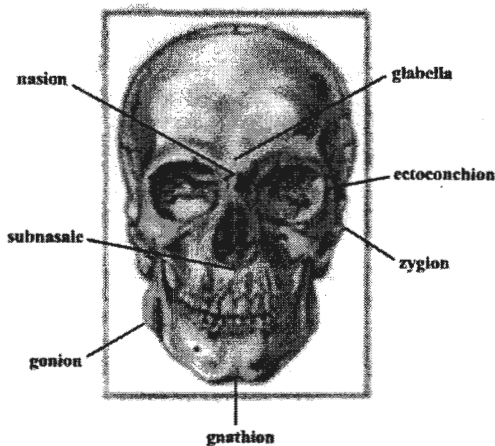
Figure 4. Diagram of Cranio-metric Points Commonly Used in Craniofacial Superimposition (Adapted from Burns 1999)

The original methodology for superimposition of handwriting employed photography to prepare the samples for presentation and transparency sheets placed over a "transmitted light table" to examine a disputed sample, superimposed over a standard sample (Osborn 1929, p. 96). A more recent methodology that seems little changed from earlier photographic methods superimposed a snapshot of a forged rubber-stamped document over an original (Hilton 1982). Both samples can be enlarged or reduced to the same size and can then be photocopied onto transparency sheets and placed on an overhead projector. In the case of handwritten documents, the forgery should match more exactly than the standard with a halting or unnatural penmanship (Slyter 1995, Hilton 1982). Yet, in the case of typewritten or rubber stamped documents the sample will show a difference in what should be identical writing (Hilton 1982). The term "standard" refers to a collection of genuine documents written by an individual against which disputed documents are measured to ascertain authenticity (Hilton 1982). This type of superimposition can also benefit from digitization, in that the samples can be scanned in and digitally manipulated for size and rotation similarity. Using the same type of editing software mentioned above, the samples can then be layered in order to compare the characteristics of each; digital superimposition is illustrated in Figure 5 below.

Figure 5: Superimposition of Handwritten "s" Image (from Brugnatelli 1999)

The criteria for analyzing questioned documents is not limited to letter formation, but includes characteristics of "writing movement," such as flow, precision, and control (Hilton 1982, p. 154). Letters themselves are subject to many kinds of variations, such as size, shape, and slant, that require the inspection of multiple documents, usually an entire standard from the writer, to be certain of letter forms unless the writer has extremely unusual habits. An excellent example can be found in the handwriting of Leonardo daVinci, in which the letters are actually backward for the most part and written from right to left (de Lucia and Borrelli 2000). Flow relates to the smoothness of pen movements, while precision refers to factors that affect legibility, like retouching. Control refers to whether the flow of the pen is continued at the end of sentences. Spacing of letters and words is also a concern. The following example, Figure 5, shows differences in the pressure used to write. A felt tip pen was used to write the letter "s" in both examples. The letter "s" from a questioned document was superimposed over a letter from a standard sample. This is only a minute portion of each sample, but several such comparisons could lead a document examiner to exclude or ID the individual who submitted the standard sample. The foreground layer also appears neater than the background layer, indicating that the time constraints for

the writing were different. This type of variation indicates that more time was taken in the writing of the former.

Dactylography has not yet benefited from the use of fingerprint superimposition as far as the current researcher is aware, so the methodology for superimposition of prints was synthesized from methods used in photographic craniofacial and handwriting models. For this study a sample of seven male and three female volunteers in Tarrant County, Texas, was taken to test the null hypothesis. All volunteers were over age 18; however, age is not a defining characteristic of fingerprints (Olsen 1978, Saferstein 1997) and this variable was not considered. Variables considered important in the sample are listed in Columns 2, 3, and 4 of Table 1. These included sex, hand (right or left), and pattern area (full or partial). A full pattern area was defined as one that included a delta and type lines, which are explained below. However, a full pattern is not the same as a full print because only the pattern area was considered for this experiment. First, thumbprints were taken from each hand and placed on half of a comic backing board (a thin sheet of cardboard) that was divided into quadrants. The upper and lower left quadrants were reserved for prints from the right hand, and prints from the left hand were placed in the upper and lower right quadrants. The researcher delineated the pattern types of the thumbprints used in the analysis, as is the common practice in dactylography (Chapel 1941, Olsen 1978, *The Science of Fingerprinting* 1985). The general pattern types are identified by the FBI (1985) as *arches*, *loops*, and *whorls*; however, since the sample of ten sets of thumbprints do not include the first pattern, the researcher will consider only *loops* and *whorls* in this study

FPC#	Sex	Hand	Area	B	RE	Total
1	M	R	P	5	6	11
2	M	R	F	4	8	12
3	F	R	P	7	4	11
4	M	R	F	6	13	19
5	M	R	P	5	7	12
6	M	R	P	10	4	14
7	F	R	P	7	9	16
8	F	L	P	3	3	6
9	M	R	F	5	6	11
10	M	L	F	8	5	13

Table 1. Sample Data and Counts from Fingerprint Analysis

Furthermore, according to the FBI (1985), the paramount concern in loops and whorls is the pattern area, which is defined as the area of *focal points*, and ridges. *Type lines* that diverge to completely surround the pattern encase this region. Focal points include the core, or estimated center of the print, and the delta, which is "in front of and nearest the center of the divergence of the type lines" (p.9). This is the region on which the study focused and identified points of similarity for comparison. This knowledge is helpful in providing a basis of orientation for superimposed prints. These areas are marked in Figure 6, along with examples of minutiae.
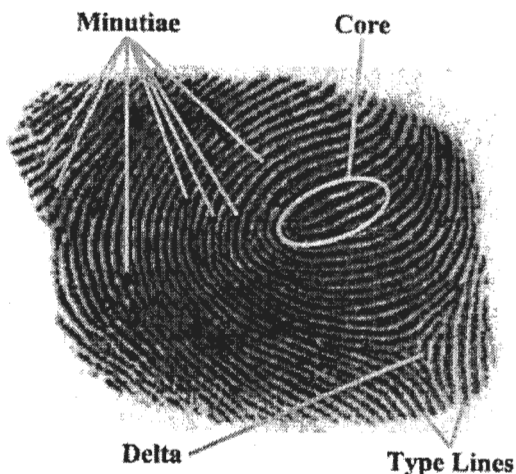
Figure 6. Diagram of a Fingerprint
(Adapted from *Science of Fingerprinting* 1985)

The prints were oriented with the type lines facing downward in prints that show full pattern areas to create a vertical axis in the upper quadrant print, which was designated Layer 0. Only a portion of the pattern area is available in the majority of sample prints, as is the case with the majority of crime scene prints (Saferstein 1997), and in six cases there is no delta present to indicate the type lines. In these latter instances the prints were oriented with the core facing upward to create the vertical axis, as Figure 7 depicts. The angle of rotation for the second print was then calculated using a Fiskars swing bar protractor and rotated by degree to match the first in orientation. The second print was then superimposed and designated Layer 1. The prints were then carefully marked points for comparison in red on the pattern area of Layer 0 and examined both layers simultaneously, setting Layer 0 at 100 percent opacity and Layer 1 at 40 percent opacity to determine how well the minutiae from the upper quadrant print matched those of the lower. All samples, except the case example from the Tarrant County Medical Examiner's office (Figure 7) were used, providing an even number of participants for the study.

The actual features used to match fingerprints, referred to as minutiae, are numerous (Galton 1892, Olsen 1978, *The Science of Fingerprinting* 1985). The research focused on bifurcations and ridge endings because these are the minutiae considered in the FBI's matching system, AFIS (Bailey personal communication). According to the FBI (1985), bifurcations are the splitting of one ridge into two ridges, and ridge endings are the points at which ridges terminate. The comparison point chosen for a bifurcation was the precise point at which the single ridge diverged. These minutiae were counted and listed in Table 1, Columns 5 ("B") and 6 ("RE"), and diagramed in the sample print shown in Figure 7. Bifurcations were abbreviated B and ridge endings were abbreviated RE in both the table and figure. To establish how the matching criteria were developed, Galton (1892) asserted that twelve points of comparison are needed to assert a positive ID; yet notable authorities on the subject do not currently require a minimum number of comparison points to ID an individual (Olsen 1978, *The Science of Fingerprinting* 1985, Saferstein 1997). According to the FBI (1985, p. 103) "the number of identical characteristics is left to the discretion of the

individual but he/she should be absolutely certain that the prints are identical before treating them as such." The research used six points from the superimposition as criteria prior to analyzing the data, due to the fact that the pattern area is often only a small portion of the actual print. More points can be used provided that the full print is legible.
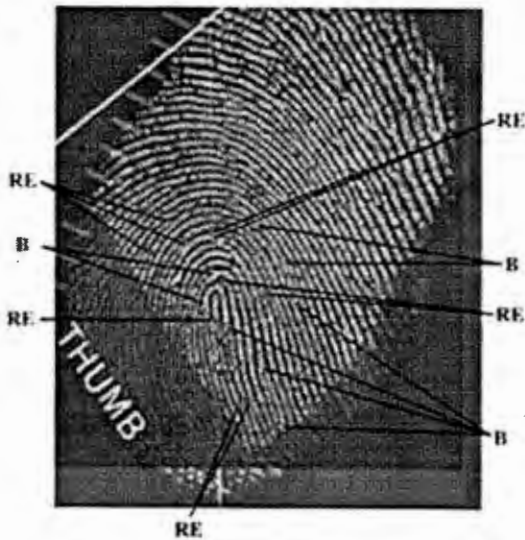

Figure 7. Superimposed Thumbprint
(Example Case from Tarrant County Medical Examiner)

STATISTICAL ANALYSIS AND RESULTS

The total matching bifurcations and ridge endings, or points of comparison, were counted in the superimposed comparisons and noted in Table 1, Column 7 (Total). The one sample t-test for means with nine degrees of freedom was chosen to analyze the data. Due to the small sample size (n=10) and the fact that the population standard deviation was unknown, a one sample t-test was preferable to a one sample z-test, which requires a large population and known standard deviation (Freedman et al. 1998). Taking into account that an average of three bifurcations and three ridge endings satisfied the requirements, a separate statistical analysis was run so that the statistical significance of each, as shown in Table 2, could be compared to the total significance of the six points used for the final analysis in Table 3. After entering the data listed in Table 1 into the computer using SPSS version 8.0, the counts were calculated by SPSS using a mean test value of three for bifurcations and ridge endings and a mean test value of six for total minutiae. These test values were used because these were the critical values established by the researcher. The resulting test statistics for bifurcations and ridge endings were 4.617 and 3.748, respectively. Both values surpassed the test statistic of 3.25 required for 0.5 percent confidence level (Freedman et al. 1998), yielding significance levels of .001 percent and .005 percent, respectively. The results for this test are noted in Table 2. These preliminary findings indicate that a match was obtained with 99.999 percent accuracy for bifurcations and with 99.995 percent confidence for ridge endings. The test calculation of total minutiae compared was even more emphatic, yielding a t-value of 5.975 and a significance level of .000 percent, which indicates nearly 100 percent confidence. This final value well exceeds the critical value and the results,

shown in Table 3. Thus, the research allows one to safely reject the null hypothesis and conclude that the dactylographic study was a success.

Bifurcations/Ridge Endings			
Test Value = 3			
t	df	Sig (2-tail)	Mean Diff
4.617	9	.001	3.00
3.748	9	.005	3.50
N	Mean	Std Dev	Std Error
10	6.00	2.05	0.65
10	6.5	2.95	0.93

Table 2. Results from First Analysis of Bifurcations and
Ridge Endings, Respectively

Total Minutiae			
Test Value = 6			
t	df	Sig (2-tail)	Mean Diff
5.98	9	.000	6.50
N	Mean	Std Dev	Std Error
10	12.50	3.44	1.09

Table 3. Results from Final Analysis of Combined Minutiae

DISCUSSION

For craniofacial superimpositions, it is recommended that one approach the problem by following the two phases of investigation enumerated by Bajnóczky and Királyfalvi (1995). First, a forensic artist must make a reconstruction of the face from a cast of the skull if a tentative lead is not already available. Photographs of the reconstruction should be made available to both the media and law enforcement for the solicitation of a tentative ID. Once a tentative ID has been established, a forensic scientist or other medicolegal professional can use the indices described by Lan and Cai (1993) to carefully photograph and superimpose the skull onto the facial photograph(s). Several points and features can then be employed in the comparison (İşcan 1993b, Burns 1999). When researchers use facial photographs of two or more views, the superimposition alone may be sufficient to make a positive ID (Austin-Smith and Maples 1994, Lan 1995). However, in cases of a single photograph, more than 13 indices and other corroborating evidence are required (Glaister 1973, Yoshino et al. 1995). The scientific certainty of the superimposition can be improved by using an antemortem radiograph compared to a skull photograph, which precludes the need for tissue depth analysis (Grüner 1993). Both craniofacial and questioned document superimpositions have been used in court to identify both victims and criminals (Nickerson et al. 1991, Hilton 1982).

Although methods of superimposition do not necessarily require an expert level of knowledge in forensic science, the professional who undertakes a project in superimposed comparison must have basic knowledge of the type of analysis needed and a clear research design prior to beginning the comparison. For example, a law enforcement officer who wishes to compare the fingerprint of an arrested suspect with a crime scene print can do so

easily by referring to *The Science of Fingerprints* (1985) for basic knowledge and following a methodology similar to that outlined above. Forensic science, like any other type of scientific endeavor, must seek to eliminate, or at least minimize Type I (false exclusion) errors. However, because of the precision needed to make a positive ID, Type II (false inclusion) errors, are no more acceptable and must also be avoided. A false exclusion for a victim would mean the misidentification of an individual, and the impact on a victim's family would be devastating. Yet, a false inclusion could lead to a lawsuit if the missing person was subsequently found. On the other hand, a false exclusion for a perpetrator could result in the release of a guilty criminal, while a false inclusion could possibly send an innocent person to jail for life, or even to a death sentence in certain states.

The utility of dactylographic superimpositions to rapidly ID or exclude individuals has been asserted from a limited preliminary experiment by in this study. The methodology used to digitally superimpose prints using pattern areas appears scientifically reliable and sound. However, a much larger sample would be needed to be completely certain of the validity of the findings, and a sample size of well over 100 individuals is recommended. It is also recommended that, with fewer time constraints, the entire print can be compared to find a mean that can be used as a guideline for the number of minutiae to be used in fingerprint comparison methods. This latter recommendation may require an even larger sample. The use of fingerprint superimposition would also enhance the presentation of a comparison in court by making the matching minutiae much easier to see. It is further asserted that the technique of superimposed comparison can be applied to many, if not all, types of comparative analysis.

CONCLUSIONS

In conclusion, one may safely infer from the preliminary study that the digital superimposition of thumbprints is more than 99 percent accurate. It must be noted that all ten superimpositions satisfied the criteria set forth of six total comparison points, and only one case was considered too illegible to show more than the minimum. The mean number of bifurcations and ridge endings were 6.00 and 6.50, respectively, and each could be matched with over 99.99 percent accuracy. If the standard error of each were subtracted from the t-values, the findings would still show greater than 99 percent accuracy for bifurcations and approximately 99 percent accuracy for ridge endings. The mean number of total minutiae was 12.50, and the prints can be matched with 99.9999, or nearly 100 percent accuracy. If the standard error was subtracted from this t-value, the result is still considerably above 99.5 percent accurate. (For comparison, thumbprints from a father and son were also superimposed to test the strength of the technique, and none of the minutiae matched exactly, as did the prints from the same person.)

ACKNOWLEDGEMENTS

made up the core of this study, enabling me to perform the statistical analyses described herein.

REFERENCES

Aulsebrook,W. A., I_can, M. Y., Slabbert, J. H., & Becker, P. (1995). Superimposition and reconstruction in forensic facial identification: A survey. *Forensic science international, 75,* 101-120.

Austin-Smith, D., & Maples, W. (1994). The reliability of skull/photograph superimposition in individual identification. *Journal of forensic sciences 39,* 446-455.

Bajnóczky, I., & Királyfalvi, L. (1995). A new approach to computer-aided comparison of skull and photograph. *International journal of legal medicine, 108,* 157-161.

Brugnatelli, F. (1999). Graphic comparisons. *Ramsey case: The documentation of the ransom note [in english].* <http://web.tiscalinet.it/faustobrugnatelli/1.htm> (1999, July 4)

Burns, K. (1999). *Forensic anthropology training manual.* Upper Saddle River, NJ: Prentice Hall. (Figure 4. Original sketch by Joanna Wallington)

Caldwell, P. C. (1986). New questions (and some answers) on the facial reproduction techniques. *Forensic osteology: Advances in the identification of human remains* (pp. 229-255). Springfield, IL: Charles C. Thomas.

Chapel, C. E. (1941). *Fingerprinting—A manual of identification.* New York: Coward McCann.

De Lucia, P., & Borrelli, G. (2000). Leonardo daVinci. *Graphology.* <http:// space.tin.it.scienza/pwvbo> (2000 June 26)

Foote, R. D. (1974). *Fingerprint identification: A survey of present technology, automated applications and potential for future development.* Huntsville, TX: Institute of Contemporary Corrections and the Behavioral Sciences, Sam Houston State University.

Freedman, D., Pisani, R., & Purves, R. (1998). *Statistics* (3rd ed.). New York: W. W. Norton

Galton, F. (1892). *Finger prints.* London: MacMillan

German, E. (2000). The history of fingerprints. *Latent print examination.* <http:// onin.com/fp/fphistory.html> (2000, May 2)

Glaister, J. (1973). *Medical jurisprudence and toxicology* (13th ed.). London: Butler & Tanner.

Grüner, O. (1993). Identification of skulls: a historical review and Practical applications. *Forensic analysis of the skull* (pp. 29-45). New York: Wiley-Liss.

Hargreaves, G., & Wilson, P. (1991). *A dictionary of graphology: the a-z of your personality* (pp. 7-8). London: Peter Owen.

Hilton, O. (1982). *Scientific examination of questioned documents.* (revised ed.). New York: Elsevier Science.

Işcan, M. Y. (1993a). Craniofacial image analysis and reconstruction. *Forensic analysis of the skull* (pp. 1-9). New York: Wiley-Liss.

Işcan, M. Y. (1993b). Introduction to techniques for photographic comparison: potential and problems. *Forensic analysis of the skull* (pp. 57-70). New York: Wiley-Liss.

Krogman, W. M., & I_can, M. Y. (1986). *The human skeleton in forensic medicine.* Springfield, IL: Charles C. Thomas

Lan, Y. (1995). A study on national differences in identification standards for Chinese skull-image superimposition. *Forensic science international, 74,* 135-153.

Lan,Y., & Cai, D. (1993). Technical advances in skull-to-photo superimposition. *Forensic analysis of the skull* (pp. 119-129). New York: Wiley-Liss.

Nickell, J. (1996). *Detecting forgery: forensic investigation of documents.* Lexington: University of Kentucky Press

Nickerson, B. A., Fitzhorn, P. A., Koch, S. K., & Charney, M. (1991). A methodology for near-optimal computational superimposition of two-dimensional digital facial photographs and three-dimensional cranial surface meshes. *Journal of forensic sciences, 36,* 480-500.

Olsen, R. D. (1978). *Scott's fingerprint mechanics.* Springfield, IL: Charles C. Thomas.

Osborn, A. S. (1929). *Questioned documents* (2nd ed.). Albany: Boyd Printing.

Saferstein, R. (1997). *Forensic taphonomy: the postmortem fate of human remains.* (pp. 437-450). Boca Raton: CRC Press

The science of fingerprinting: classification and uses. (1985). Washington, DC: U.S. Dept. of Justice, Federal Bureau of Investigation. For sale by the Superintendent of Documents, U.S. Government Printing Office

Slyter, S. A. (1995). *Forensic Signature Examination.* Springfield, IL: Charles C. Thomas

Yoshino M., Imaizumi, K., Miyasaka, S., & Seta, S. (1995). Evaluation of anatomical consistency in cranio-facial superimposition images. *Journal of forensic sciences, 74,* 125-134.

Yoshino, M., Matsuda, H., Kubota, S., Imaizumi, K., Miyasaka S., & Seta, S. (1997). Computer-assisted skull identification system using video superimposition. *Forensic science international, 90,* 231-244.

Beethoven's Orchestration of the Trombone in His Fifth Symphony and Its Importance to the History of the Instrument

Joseph Rodriguez
Mentor: Darren McHenry, Department of Music

ABSTRACT

Beethoven was an influential composer who helped bridge the gap between the Classical and Romantic Periods. He also played a significant role in the history of the trombone by using it in his Fifth Symphony. By examining the history of the trombone up to the time of Beethoven and analyzing his use of the instrument, it is possible to reveal the impact Beethoven had on the future use of the trombone in symphonies.

INTRODUCTION

Have you ever listened to a symphony and thought, "Wow! I wonder how the composer created that incredible sound that sent chills up my back?" This paper will provide some insight as to how Beethoven used the trombone to create a triumphal sound in the last movement of his Fifth Symphony. It will also discuss the importance and influence of Beethoven's Fifth Symphony on the development of the trombone. First, it is important to trace the history of the trombone up to the time of Beethoven.

Early History of the Trombone

Historians do not know many details of the history of the trombone due to the lack of written descriptions, instruments from the period, and, of course, sound recordings of these instruments. Some estimations and assumptions have been made regarding the trombone's origins because of this lack of information, such as the theory that the trombone more than likely evolved from the trumpet:

> In fact the name *trombone* comes from the Italian word *tromba* meaning 'big trumpet.' The German term *posaune* can also be traced to the trumpet, having derived from the *busine* or *buzine*, the medieval straight trumpet. The old English appellation *sackbut* came from the Spanish name *sacabuche*, a pump or draw-pipe. The French term *sacqueboutte* likely has its origin from the same source (Nicholson 70).

The slide was an invention that changed the way a brass instrument could be played. Before the invention of the slide, each brass instrument could play only the notes in the overtone series. According to Baines, the slide was first used on trumpets in Europe sometime in the early fifteenth century (94). The new invention led instrument makers to create slide trumpets in lower pitches; it was such low-pitched slide trumpets that came to be known as trombones. When the slide was fully developed, the instrument could produce all twelve chromatic pitches. Therefore, the trombone was the first fully chromatic brass instrument. As Nicholson states, "It was perhaps the first of all the orchestral instruments to arrive in its present shape" (70). Even though there is not enough information to know the true birth date of the trombone, most historians believe the trombone probably appeared as an instrument in the middle of the fifteenth century (Baines 107).

In Georg Neuschel's letters, dated around 1550, he referred to "two pairs of *Bogen* for the *Quartposaune*" (Baines 116). A *Quartposaune* was a bass trombone in the key of F and the pairs of *Bogen* are what we know as crooks. A crook was a short, curved piece of extra tubing that was connected at the top curve between the bell and the slide of the trom-

bone in order to lower the pitch of the instrument, sometimes up to a fourth or fifth (see Figure 1). Neuschel's letter is the earliest source providing information on the use of crooks. Therefore, as far as we know, the trombone was the first brass instrument to use crooks (Baines 116). In the seventeenth century the family of trombones was made up of the alto, tenor, bass, and contrabass trombones. Praetorius in his *De Organographia*, dated 1618, described four kinds of slide trombones of the time:
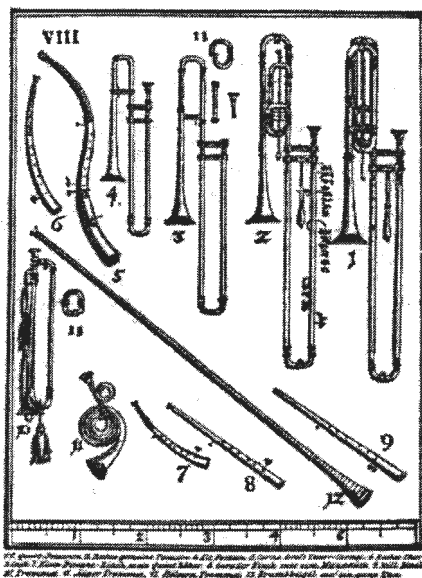

Figure 1: Praetorius, *De Organographia* (Syntagma II), 1618, Pl. VIII

Alto	*Alt* or *Discant Posaun* in F, Eb and D.
Tenor	*Rechte Gemeine* (Ordinary) *Posaun* in Bb, A, E and D.
Bass	*Quart* or *Quint Posaun* in F and Eb respectively.
Contrabass	*Octav Posaun* in BBb or an octave below the Tenor.

Table 1: Description of Original Names and Most Common Keys of Trombones in Praetorius' *De Organographia* (Nicholson 71).

The trombone was used in various ways in different musical periods. During the Renaissance, the trombone was among the most popular instruments. Wealthy families would hire trombonists to play in their private ensembles. Trombonists would play in all sorts of public and private gatherings, including weddings, funerals, banquets, processions, town band concerts, and some early operas and dramatic productions. The trombone also had an important role in church music. Churches would hire trombonists to accompany the choir or to play in different ensemble features. The Moravian Church had a tradition of using a trombone choir to accompany the church in worship.

One popular composer for brass instruments during the Renaissance was Giovanni Gabrieli. He was one of the first composers to assign certain instruments to the different parts in his music. Before Gabrieli, composers left the choice of instrumentation up to the performer. He was also one of the primary composers to write in an antiphonal style. An-

tiphonal style is an approach in which two to three choirs of instruments or vocalists are placed in separate areas of the church or other performance venue. The antiphonal style of writing was very popular during this period and has continued to heavily influence the way composers write music today. Another popular composer for brass instruments during the Renaissance was Daniel Speer. He wrote several sonatas for three trombones and continuo. Gabrieli and Speer were very influential Renaissance composers.

During the Baroque Period, the use of the trombone declined. J. S. Bach was an important composer during this time, but he wrote for the trombone in only fifteen out of his two hundred and sixteen cantatas. Overall, there was little Baroque music written for the trombone. After the Baroque, came the Classical Period, which included composers such as Gluck, Haydn, Beethoven, and Mozart. Trombones were used primarily in church music, operas, oratorios, and masses. Composers also started to write for trombones grouped as a three-member section. Mozart used three trombones in *The Magic Flute*, *Don Giovanni*, and the *Requiem*, three works that had a significant impact on the future of the trombone. For example, Mozart's *Requiem* has a trombone solo in the *Tuba Mirum* section of the work, which to this day remains a major solo in the trombone repertoire.

Beethoven was a unique composer whose music has been classified in both the Classical and Romantic Periods, due to the fact that his writing style changed over the years. In his earlier years he wrote in a Classical style similar to Haydn, but as time went on, he developed a more dramatic or heroic style. His Fifth Symphony was written in the latter style. In fact, most scholars consider Beethoven as an inventive composer who bridged the gap between the Classical and the Romantic Periods.

BEETHOVEN'S INFLUENCE

Beethoven was also an ingenious composer whose music has been well respected by audiences, performers, and other composers for nearly two hundred years. Franz Liszt wrote a letter in 1852 to Wilhelm von Lenz in which he stated, "To us musicians the work of Beethoven parallels the pillars of smoke and fire which led the Israelites through the desert" (Crofton 16). Friedrich Nietzsche wrote in 1878, "Beethoven's music is music about music" (Crofton 16). Beethoven's Fifth Symphony is one such masterpiece. This symphony was first performed on December 22, 1808, in the Imperial, a private theatre at Vienna. Other pieces performed in the same concert were the *Chorale Fantasia*, the premier of the *Pastoral Symphony* (6th), and the *Piano Concerto No. 4* in G Major, which Beethoven himself performed (Wier 132, 167). In his *Memoirs* (1865), Hector Berlioz quoted Jean François Le Sueur, speaking of Beethoven's Fifth Symphony, "Ouf! Let me get out; I must have air. It's incredible! Marvellous! It has so upset and bewildered me that when I wanted to put on my hat, *I couldn't find my head…*One ought not to write music like that" (Crofton 16).

"Fate knocking at the door" is the phrase that Beethoven supposedly used to describe the opening theme of the first movement (Figure 2). Except for this coined phrase, there is no other written indication that the Fifth Symphony tells a story. Although Beethoven never expressed the Fifth Symphony in words, his music tells a dramatic story of conflict with a triumphal conclusion. Some historians and musicologists have used their imagination to describe its story in words. One such analyst, who had an incredible imagination, was, Hector Berlioz, who described all nine of Beethoven's symphonies in his book, translated by De Sola, entitled *Beethoven: A Critical Appreciation Of Beethoven's Nine Symphonies and His Only Opera, Fidelio, With its Four Overtures*. In his description of the Fifth Symphony, he

describes the first movement as "the terrible fury of Othello when hearing from Iago's mouth the poisoned calumnies persuading him of Desdemona's crime" (28). Tovey thinks that the first movement is "in a mood of Shakespearian humour [sic]" (41). Downes contrasts these pictures of the first movement by envisioning a poem, where a head peeks over a wall where there are people having fun on the other side. Then the head, which is the "prophet of war," makes a "horrible laugh" (62). The second movement is mysterious, in some parts foreshadowing the triumph of the fourth movement.

Figure 2: Opening Theme to Beethoven's Fifth Symphony

Berlioz's description of the transition from the third to the fourth movements is descriptive, imaginative, and effective:

> The Scherzo (the third movement, marked *Allegro*) is a strange composition; its first bars, although presenting nothing terrible, cause the strange emotion we experience under the magnetic glance of certain individuals. Everything in it is mysterious and somber. The orchestral devices appear somewhat sinister, appearing to belong to the same order of ideas creating the famous Bloksberg scene in Goethe's *Faust*.

> *Piano* and *mezzoforte* tints prevail. The middle part, or trio, is remarkable for a brass passage executed with all the force of the bow. Its uncouth weight shakes the feet of the player's music stands, and somewhat resembles the frolicking of a jolly elephant. But the monster departs, and the noise of his mad careening dies away gradually. The scherzo motive reappears pizzicato; peace is regained gradually until nothing more is heard than a few notes plucked daintily by the violins, plus the strange clucking of the bassoons giving their high A flat closely opposed by G as octave in the chord of the dominant minor ninth.

> Then the strings, col arco, interrupting the cadence, softly take the A-flat chord as their resting place for a bit. The rhythm depends entirely on the kettledrum, and is sustained by light strokes struck by sponge-covered sticks, forming a dull design against the stagnation of the rest of the orchestra.

> The kettledrum note is C, whereas the key of the movement is C minor; but the A-flat chord, long sustained by other instruments, seems on one hand to introduce a different tonality, while on the other the isolated martellato of the kettledrum on C tends to preserve the spirit of the original key.

> The ear hesitates, uncertain as to how this harmonic mystery is about to resolve itself, when the dull pulsations of the kettledrum, becoming more and more intense, meet the violins who have rejoined the rhythmic movement and changed the harmony. The chord is now the dominant seventh (G, B, D, F) where the kettledrum stubbornly continues its roll upon the C tonic.

> And then the entire orchestra, reinforced by the trombones which have not appeared before, bursts forth in the major mode on a triumphal march theme, and the finale begins [sic] (29).

It is apparent that some kind of conflict is being powerfully resolved in the fourth movement.

BEETHOVEN'S CHOICE OF INSTRUMENTATION

Having discussed the dramatic sequence of the Fifth Symphony, it is important to examine the fourth movement in order to discuss some of the tools, primarily the use of the trombone, which Beethoven used to achieve the dramatic and joyful ending. Perhaps the most obvious is the change of feeling. Throughout the symphony the primary key is C minor until the fourth movement where the tonic stays the same but the mode changes to C major. Music in the minor mode tends to create a sad or mysterious mood. In contrast, music in the major mode tends to create a cheerful mood. The change of mode helps give the fourth movement a rejoicing spirit filled with happiness. The style also changes in the opening of the fourth movement. Compared to the fury of the first movement, the drama of the second movement, and the mysteriousness of the third movement, the opening of the fourth movement resembles a victorious march. The tempo is marked at half note equals 84, which indicates a cut time feel such as in a march. This change in style is another way in which Beethoven creates a sense of victory in the fourth movement.

Beethoven also expands the orchestra as another tool to enhance the impact of the opening measures of the fourth movement. Up to this point the general size of the orchestra consisted of full strings, two timpani, two horns, two trumpets and eight woodwind instruments (which include two flutes, two oboes, two clarinets and two bassoons). In his Third Symphony, Beethoven added a third horn to this standard instrumentation. However, in his Fourth Symphony he deviated from his standard instrumentation by using only one flute, making a seven member woodwind section. In the Fifth Symphony, Beethoven enlarged the orchestra considerably by using two extra woodwinds (a piccolo flute and a contrabassoon), and three extra brass instruments (an alto, a tenor, and a bass trombone). Even though these instruments are used in the Fifth Symphony, they are not played in all of the movements. As a matter of fact, they only play in the fourth movement. By saving this expansion of the orchestra until the last movement, Beethoven is able to create a more dramatic effect by providing more sound.

The choice of instrumentation is also an indication of Beethoven's intent to create a larger, more dramatic sound in the fourth movement. It is necessary to note the pitch of the extra instruments used in this movement: the piccolo flute (the highest pitched instrument of the woodwind family), the contrabassoon (the lowest pitched instrument of the woodwind section), and the trombone section, consisting of alto, tenor, and bass trombones (pitched in the high, middle, and low ranges of the brass section, respectively). By choosing instruments of extreme ranges, such as the piccolo flute and the contrabassoon, Beethoven increased the range of the orchestra, thereby creating a new timbre. By using the trombones, he added power and depth to the orchestral sound. Trombones add power because of the acoustics of the instrument: they have the capability of playing loud. Trombones add depth because their range lies in the middle-to-low range of the orchestra. Beethoven's choice of instrumentation has great influence on the sound of the fourth movement.

Having discussed some of the different tools used to create the sound of the fourth movement, let us now look in more detail at how Beethoven utilizes the trombone in his Fifth Symphony. To analyze Beethoven's use of the trombone, it is necessary to group measures and then make generalizations on how the role of the trombone compliments

the music being played in that movement. Table 2 provides a road map of every measure in which the trombones play and provides an analysis of their role in the orchestration.

Measures	Dynamic	Orchestration	Role of Trombones	Place in Form
1-4	ff	Tutti	Melody doubled (inverted)	First Theme
5-42	ff	Tutti	Provide pulse and attack	
49-63	f	Tutti	Provide sonority, pulse, and attack	Second Theme
69-71	p	w.w. and strings	Provide pulse and sonority	Codetta
72-81	f	Tutti	Melody doubled	
82-90	f	Tutti	Provide pulse and attack	
112-118	f	w.w. and strings	Melody	Development
120-153	f	Tutti	Provide sonority and pulse	Retransition (m. 132)
207-269	ff	same as 1-63	same as 1-63	Recapitulation
281-286	f	Tutti	Melody doubled	Codetta
286-302	f	Tutti	Provide sonority and pulse	
302-307	f	Tutti	Melody doubled	Coda
308-311	f	Tutti	Provide sonority	
312-349	f	Tutti	Provide attack	
386-390	f	Tutti	Provide sonority and crescendo	(Presto m. 362)
411-444	ff	Tutti (last to enter)	Provide attack	

**Table 2: The Role of the Trombone in
Beethoven's *Symphony No. 5* in C minor – IV. Allegro**

Many key points need to be discussed regarding Beethoven's scoring for the trombone in his Fifth Symphony. Let us first look at the opening statement in the fourth movement. The alto trombone reinforces the melody by playing it in an inversion. Beethoven gives the alto trombone this form for two reasons: he wants it to reinforce the overall melody, and he wants the impact of the high C on the first note. The trombones have melodic material five times throughout the last movement, but they are the primary melodic instruments only once in measures 112-118. In these measures, the melody is a development of the theme that occurs at the beginning of the movement. This line is very interesting because it has been played several times earlier in the piece: by the cellos in measures 45-46, 50-54; by the cellos and violas in the development section in measures 91-95; and by the cellos, basses and contrabassoon in measures 106-112. When the trombones enter in measure 112, it is the first time this line is introduced as the melody. One way Beethoven indicates this is by scoring both alto and tenor trombones on the same pitch marked *forte* while doubled by the two bassoons an octave lower marked *fortissimo*.

Beethoven wrote fairly independent lines for the trombone. One example occurs in measures 69-71, where the trombones are the only brass instruments accompanying the woodwinds' melody. Another instance is in measure 130, where the alto and tenor trom-

bones are the only instruments that have sustained notes, helping to build to the climax in measure 132. In measures 267, 268, and 308-312, the trombones are the only instruments to have long notes, while the rest of the orchestra has fast notes. One last example where the trombones have an independent line occurs in measures 385-390. Here, complimented by the piccolo flute, the trombones have long cascading notes as they build into the last major statement of the theme in measure 390.

It is also interesting to note that the rhythm of the trombone parts is generally simplified compared to the rhythms of other instruments of the orchestra. Throughout the whole fourth movement the trombones never have a rhythm faster than the equivalent of a quarter note. However, in the third measure the trombones have eighth notes separated by eighth rests. This rhythm corresponds to quarter notes or one note per beat (see Appendix 2). In measures 4 and 5, the trombones are the only instruments that have a rest on the fourth beat instead of the dotted eighth-sixteenth note rhythm included in most of the other parts. Another example of simplified rhythm is apparent in measure 17 where the whole orchestra has dotted eighth-sixteenth note rhythms while the trombones have quarter notes.

Beethoven simplified the rhythm, but he stretched the range of the alto trombone part. The ranges for the three trombone parts in the Fifth Symphony are given in Figure 3 below:
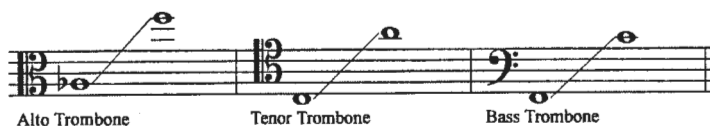
Alto Trombone   Tenor Trombone   Bass Trombone
Figure 3: Ranges of the Trombone Parts in Beethoven's Fifth Symphony

In measures 80 and 81 the alto trombone has the dreaded high F (an octave and a fourth above middle C) held for two tied whole notes. Also near the end of the piece, in measures 430 and 431, the alto has another tied whole note rhythm, this time on a high E (an octave and a third above middle C). Throughout the movement the alto trombone stays primarily above middle C. The high F in measures 80 and 81 is still one of the highest, most demanding notes written for trombone in all orchestral literature. This is significant because it shows the boldness of Beethoven's orchestration of his first symphony with trombone. He wrote only up to a high D (a ninth above middle C) in his Ninth Symphony.

Dynamics play a key role in Beethoven's orchestration of the trombone. In general, when he wanted it loud he added the trombones. Beethoven generally wrote block dynamics, meaning that he wrote the same dynamics on all parts at any given point. There are a few exceptions, however, dealing specifically with the trombone. For example, in measure 58 the whole orchestra is *fortissimo* except for the trombones that are marked only *forte*. Also in measure 112, where the trombones have the melody doubled by the bassoons, the trombones are marked *forte* while the bassoons are marked *fortissimo*. Out of the 215 measures in which the trombone plays, 212 measures are marked either *forte* or *fortissimo*. This proves that Beethoven thought of the trombone as a powerful instrument, although he did not rule out its effectiveness when playing soft. Measures 69-71 are marked at the other extreme, *piano*.

Frequency of use is also an important issue when analyzing the role of the trombone in Beethoven's Fifth Symphony. The trombones play in about half of the measures of the fourth movement. Before the recapitulation, Beethoven includes the last part of the third

movement labeled *Scherzo*. In this section, Beethoven does not use the trombone, probably because he did not use it in the third movement, and he wants to create the same dramatic effect on the recapitulation as he did at the beginning of the movement. Also, with the exception of two entrances, the trombones play only when the whole orchestra plays (see Table 2). This is significant because it shows the primary role of the trombone as a tool for reinforcing the sound of the entire orchestra.

BEETHOVEN'S LEGACY

Beethoven's use of the trombone in the Fifth Symphony had a great impact on the manner in which subsequent composers orchestrated their symphonies. Two composers that modeled Beethoven's use of the trombone were Hector Berlioz and Johannes Brahms. Berlioz used trombones in the fourth and fifth movement of his *Symphonie Fantastique*. Although this piece has five movements, some scholars group the fourth and the fifth movements as one. In Brahms' *Symphony No. 1* in C Minor, he also saved the trombones for the last movement, featuring them in a glorious transition from C minor to C major for the final theme. Other major composers, such as Schubert, Mendelssohn, Tchaikovsky, Bruckner, Mahler, Dvorak, and Shostakovich, continued to develop the use of the trombone and increase its significance in their symphonies. Because of Beethoven's successful use of the trombone in his Fifth Symphony, as suggested by the popularity of the work even today, subsequent composers have continued to develop the significance of the trombone in their works.

BIBLIOGRAPHY

Baines, Anthony. Brass Instruments - Their History and Development. New York: Dover Publications, Inc., 1993.

Bate, Philip. The Trumpet and Trombone: An Outline of their History, Development and Construction. London: Ernest Benn Limited, 1966.

Berlioz, Hector. Beethoven: A Critical Appreciation of Beethoven's Nine Symphonies and His Only Opera, Fidelio, With Its Four Overtures. Trans. Ralph De Sola. Boston: Crescendo Publishing Co., 1975.

Burrows, Raymond. Symphony Themes. New York: Simon and Schuster, 1942.

Crofton, Ian, and Donald Fraser. A Dictionary of Musical Quotations. New York: Schirmer Books, 1985.

Downes, Olin. Symphonic Masterpieces. New York: Tudor Publishing, 1941.

Fischer, Henry George. The Renaissance Sackbut and Its Use Today. The Metropolitan Museum of Art, 1984.

Lee, James D. A Comparative Study of the Symphonic Trombone Scoring of Beethoven, Brahms, and Berlioz. Diss. Kent State University Graduate College, 1974. Kent: Kent State University, 1974.

Nicholson, Joseph M. «The Trombone: Its Evolution and History.» Music Journal. October 1967: 70-79.

Sadie, Stanley, ed. The New Grove Dictionary of Music and Musicians. Vols. 17, 19. Washington D.C.: Groves Dictionaries of Music Inc., 1980.

Tovey, Donald Francis. Essays in Musical Analysis. Vol. 1. London: Oxford University Press, 1968.

Weir, Albert E. The Nine Symphonies of Beethoven in Score. New York: Bonanza Books, 1935.

APPENDICES

Appendix 1:
   Timeline of Beethoven's Symphonies Orchestrated with Trombones

Symphony #5	Symphony #6		Symphony #9
1807-1808	1808		1822-1824
Premiered:	Premiered:		Premiered:
Dec. 22,1808	Dec. 22,1808		May 7,1824

| 1805 | 1810 | 1815 | 1820 | 1825 |
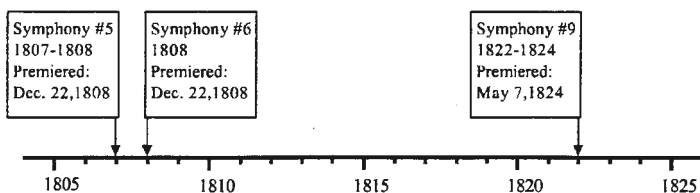
Appendix 2:

Scores to the Fourth Movement of Beethoven's Fifth Symphony (Weir 71-74)

Score to the Fourth Movement of Beethoven's Fifth Symphony (Weir 71)

Score to the Fourth Movement of Beethoven's Fifth Symphony (Weir 72)

Score to the Fourth Movement of Beethoven's Fifth Symphony (Weir 73)

Score to the Fourth Movement of Beethoven's Fifth Symphony (Weir 74)

Score to the Fourth Movement of Beethoven's Fifth Symphony (Weir 75)

Describing Postpartum: Adolescents' Experiences with Postpartum Depression, Breastfeeding, and Birth Control

Jennifer E. FitzGerald

Mentor: Dr. Cheryl Anderson, School of Nursing

ABSTRACT

A descriptive exploratory study using methodological triangulation researched postpartum depression levels, use of birth control and breastfeeding patterns of ten adolescent mothers previously enrolled in Healthy Beginnings (HB). HB is an educational program at the University of Texas at Arlington offering services prenatally and after pregnancy to low risk and teen mothers. The follow-up of the subsample of adolescents at two to eight months postpartum was a part of a larger program evaluation study contacting all previous HB enrollees. Follow-up was conducted using a semi-structured interview via telephone. Program participants were asked a series of questions regarding their experiences with postpartum depression, breastfeeding, and birth control use, generating qualitative data. Similar themes became evident among all of the teens. The Edinburgh Postnatal Depression Scale (EPDS) was also administered to teens over the telephone as an objective indicator of postpartum depression (PPD) generating quantitative data. Several adolescent mothers were found to be experiencing mild to severe PPD, with only a minority of teens disclosing past or current breastfeeding of infants or birth control use.

INTRODUCTION

The period of adolescence is a tumultuous time filled with impending developmental and physiological changes. While pregnancy is an emotionally stressful period in any woman's life, the transition from childhood to parenthood for the adolescent mother is often overwhelming. "Adolescent pregnancy and motherhood superimpose a situational crisis upon the ongoing developmental crisis of adolescence and pose many problems for both the individuals involved and society as a whole" (Chen, Chou, Tseng, & Wang, 1999 p. 459). The situational crisis of an unplanned pregnancy may have a profound adverse effect on the adolescent mother, her child, and her family.

In an attempt to provide adolescent mothers with needed maternity care and education at no cost, the obstetrics faculty at The University of Texas at Arlington, School of Nursing, developed a program pairing senior nursing students with expectant mothers as a part of their clinical rotation for graduation. Under the supervision of clinical nursing faculty, the senior obstetrical nursing students provided both prenatal and postnatal care to selected expectant mothers and their families. As a part of required assessments during home care, nursing students measured fundal height, performed Leopold maneuvers, administered urinalysis, and checked vital signs. Teens were also appraised for feelings of depression and use of breastfeeding and birth control.

LITERATURE REVIEW

Postpartum depression (PPD), which has been reported to affect 10 – 15% of new mothers (Wisner & Wheeler, 1994), is a devastating condition that has often been overlooked among the adolescent childbearing population. Common symptoms characterizing PPD can include feelings of inadequacy, anxiety, despair, lack of energy, loss of interest in sexual activities, and compulsive thoughts (Beck, 1992). In her phenomenological study of the lived experience of postpartum depression, nurse researcher Cheryl Beck engaged in conversation with adult mothers who suffered from PPD. Some mothers described the experience of postpartum depression as a "...living nightmare filled with uncontrollable anxiety attacks, consuming guilt, and obsessive thinking." (Beck, 1992, p. 166) Beck also notes

that adult mothers suffering from PPD disclosed that they had considered hurting themselves as well as their infants (1992 p. 166). In a similar study by Beck, adult mothers suffering from postpartum depression described feeling like "…robots just going through the motions of taking care of their babies" (Beck, 1993 p. 45). The manifestation of depressive symptoms can surface within the first few days after delivery in the more common, milder stage known as "baby blues," or can become worse as long as six months or more after childbirth (Beck, 1993).

Among adolescent mothers, low self-esteem prevailed as a critical risk factor in the development of postpartum depression (Chen, 1996). Adolescent mothers are also at an increased risk for developing both physical and psychological complications between mother and baby (Nichols & Zwelling, 1997). Smiling, en face looking, kissing, close holding, and stroking of a new infant are parental behaviors which are characteristic of positive maternal-infant attachment (Nichols & Zwelling, 1997) and are often absent among depressed mothers. Poor maternal-infant attachment can unfortunately result as the emotionally unstable adolescent attempts to cope with her new role and responsibilities. This maladaptive behavior may consequently result in negative long-term outcomes for the child of a teen mother, such as failure to thrive, or child neglect and abuse. Since the first few days following childbirth are a critical period for the development of mother-infant bonding, as well as the establishment of trust, it is crucial for health care providers to be able to identify those adolescent mothers who are at risk for developing PPD.

Breastfeeding, which has long been recognized for its therapeutic (enhancing attachment) and nutritional benefits, is often rejected by adolescent mothers as an option for their infant. Research has shown that breastfeeding initiation rates among adolescent mothers are generally lower than among adult mothers. In 1995, 42.8% of teen mothers initiated breastfeeding, as compared to 45% to 70% of adult mothers who initiated breastfeeding during the same time period (Ryan, 1997 as cited in Wambach & Cole, 2000). A depressed new mother may never consider breastfeeding.

While many social and economic hardships affect young adolescent mothers, these disadvantages could potentially be offset by the utilization of breast milk for infant feeding (Wambach & Cole, 2000). Adolescent mothers typically tend to utilize bottle-feeding, however, as a preferred method of infant feeding for numerous reasons. Negative attitudes about breastfeeding and its perceived disadvantages such as personal inconvenience, embarrassment, and fear of pain have been cited as reasons not to breastfeed among teen mothers (Wambach & Cole, 2000). Additionally, an adolescent mother often finds it difficult to transcend from a "self-centered focus" to an "other-centered focus" in order to fully appreciate and understand the value of breastfeeding her infant (Wambach & Cole, 2000). Teens' egocentric perspectives can prevent them from performing the unselfish act of breastfeeding.

The frequency of PPD, the common disregard for breastfeeding and the potential poor maternal-infant attachment among teen mothers and their infants, coupled with the high number of teens giving birth each year, presents a major social problem. Research conducted on recidivism rates among teens suggests that within 2 years after a first pregnancy, 40% to 50% of teen mothers will become pregnant again (Guttmacher Institute, 1994 as cited in Nichols & Zwelling, 1997). The economic burden of adolescent pregnancy affects the nation as a whole, as reflected by the more than $25 billion dollars that is spent every year on federal assistance programs which provide support to mothers who first delivered as an adolescent (Center for Population Options, 1992 as cited in Nichols & Zwelling,

1997). Although the use of condoms has increased, many teens still unfortunately engage in unprotected sex and report that their pregnancies were unintended (Rodriquez & Moore, 1995 cited in Mackey & Tiller, 1998). Many of these unplanned pregnancies may result in subsequent emotional turmoil for the adolescent mother and her family members, with neglect and abuse of an infant as an unfortunate consequence.

METHOD

Design

As part of a larger evaluative study for the HB program, a small descriptive study exploring teens' postpartum depression and use of breastfeeding and birth control was conducted using methodological triangulation. Methodological triangulation, the combination of two or more research methods, was utilized to ensure a complete understanding of the major concept of postpartum depression (Morse, 1991). Use of interviews and one survey tool for PPD generated both quantitative and qualitative data, enhancing the comprehensiveness of measurement of the variable. Identical questions, assessing PPD, breastfeeding, and birth control were asked of all teens with gentle probing for information. Data from participants was collected until several consistent themes emerged in relation to all three concepts explored.

Sampling and Subjects

A list of all previous enrollees in HB (n=39), consisting of 22 adolescent mothers and 17 adult mothers (20 years of age and older) was obtained. A systematic random sampling, with a random start, identified ten adolescent mothers after five passes for contact. The adolescents selected to participate in the study were contacted by telephone by the examiner. The sample consisted of three Caucasian, four Black, and three Hispanic adolescent females. The participants' educational levels ranged from grade 8 through grades 11. The participants' ages ranged from 16 to 18 years old with a mean age of 16.8 years. This was the first child for all ten participants. All teens agreed to participate in the study; however, the amount of information provided varied according to each teen.

DATA COLLECTION

Procedure

Data was collected via telephone using a semi-structured interview. Participants were asked open-ended questions about postpartum depression, breastfeeding, and birth control use. Gentle probing was done which led to informative "conversations" with the teens. Questions such as "What made you decide to use (not use) breastfeeding?" and "What are you reasons for deciding to use (not use) birth control?" were chosen as appropriate. Questions were asked of every teen in the same order and tone of voice to every teen by the researcher. Only one researcher, that had past experience in the HB program as a senior nursing student, collected the information. A teen mother's responses to the researcher were rephrased and repeated to the teen for confirmation and validity of information.

Instrument

In addition to the semi-structured interview, the Edinburgh Postnatal Depression Scale (EPDS) was administered to the research participants upon the conclusion of the interview. This triangulation of methods provided a more valid and reliable estimate of the teen's

depression scores. Obtaining an accurate measure of PPD among the teens was necessary, as it was assumed that depression was an influence on the teen's choice of infant feeding and birth control.

The EPDS has a high level of validity and reliability and has become widely utilized as an effective screening tool in the assessment of prenatal and postpartum depression (Clifford, Cox, Day & Werrett, 1999). Participants' responses to the Edinburgh Postnatal Depression Scale (EPDS) were scored and ranked accordingly. EPDS scores of 9 or greater can be considered indicative of postpartum depression if other variables are present. Scores of 12 or greater indicate depression and should serve as a "red flag" for health care personnel (Cox, Holden, & Sagovsky, 1987 as cited in Clifford et al., 1999).

DATA ANALYSIS

Nearly half of the teens interviewed (40%), were found to be depressed (n=4) with EPDS scores of 12 or greater. The same teens also verbalized depression. EPDS scores ranged from 1 to 24 with a mean of 8.6 and a standard deviation of 7.17. An additional teen had an EPDS score of nine; however, she was excluded from the group of teens categorized as depressed because of verbal statements to the researcher stating that she did not feel depressed. Qualitative data collected from this participant did not reveal any significant data that would place her at risk for PPD. The researcher, therefore, found that both qualitative and quantitative data coincided to present an accurate assessment of PPD among the teens being studied (Table 1).

EPDS Scores	Verbal Communication of Depression	
	Yes (n=4)	No (n=6)
1-8	(n=0)	(n=5)
9-11	(n=0)	(n=1)
12-30	(n=4)	(n=0)

Table 1. Correlation of EPDS Scores with Verbal Communication of Depression

Describing Adolescents' Experiences with Postpartum Depression

From teen "conversations" four themes emerged regarding PPD: multiple roles and responsibilities, relationship problems, self-oriented focus, and lack of resources.

Multiple Roles/Responsibilities

In describing their roles as mothers, many teens reported feeling anxious and overwhelmed by the demands of parenting. The multiple roles that teens must juggle were found to be a significant source of anxiety in the lives of teen mothers. Ms. A, a seventeen-year-old mother of a three-month-old infant stated, "I get stressed out because I'm working, watching the baby, and going to school at the same time!" Teen mothers frequently expressed that they were previously unaware of the difficulties that would accompany childrearing. Often, this newfound responsibility left them feeling inadequate in their efforts to provide care for their infants. One teen, Ms. L, a seventeen-year-old mother of a two-month-old infant stated, "When he (infant) does a lot of crying, I don't know what to do to calm him down." Feeling overwhelmed and stressed by the responsibilities of parenting,

one teen, Ms. K, an eighteen-year-old mother of twin babies stated, "I didn't think it (raising children) would be this hard…I feel overwhelmed, stressed, abandoned, and neglected…I'm to the point right now that I want some one to take them (infants)."

Relationship Problems

Half of the teen mothers surveyed (n=5) reported that they were experiencing problems in their personal relationships. Difficulties in the relationships that they shared with the baby's father appeared to be the source of many of their frustrations. Although most of the teens were reluctant to discuss specifically what problems they were dealing with in their relationships, they did disclose their dissatisfaction with the situation. Abuse (physical, emotional, and verbal) was also suspected in some of the teen relationships. Ms. J, a seventeen-year-old mother of a five-month-old infant stated, "I'm going through a lot of stuff. Stuff is going on with my baby's daddy!" Ms. K, another teen who expressed her unhappiness with her infants' father stated, "I've got a strain in my marriage. He (husband) always leaves and goes places and doesn't come home until four or five in the morning!"

Self-Oriented Focus

When discussing teens' personal reasons for feeling depressed, most viewed their problems with a self-oriented focus. Consumed with their own needs, teens were often incapable of acknowledging the needs of others, especially their infants. Unable to seek solace externally, most teens turned their thoughts inward. One teen, Ms. K, angrily stated, "Nobody cares about what I say or want…I'm always stuck with them (infants)… I never get to go anywhere!"

Lack of Resources

Teen mothers lacked the financial resources to deal with their emotional problems. Limited finances often kept teens from seeking medical treatment for PPD. While discussing treatment options for PPD, one teen mother disclosed that she had wanted to seek help for her depression, however, due to her finances, she was unable to do so. She stated, "I haven't gotten to see anyone yet. I don't have the money. I can't work and my husband can't get a job. That makes him upset." Feelings of depression can consequently prevent teens from making good choices, such as choosing breastfeeding or using birth control.

Describing Adolescents' Experiences with Breastfeeding

While 60% (n=6) of the teens said that they planned to breastfeed their infants during their prenatal assessment, only 30% of the teen mothers who were interviewed initiated breastfeeding. These mothers breastfed their infants for an average of 3.3 weeks before they decided to bottle-feed. Three themes became evident while discussing the adolescent mothers' decision to choose bottle-feeding over breastfeeding: convenience, misconceptions about breastfeeding, and improper breastfeeding techniques. Some teen mothers (n=3) felt that their decision to bottle-feed was based on the fact that bottle-feeding was more convenient than breastfeeding. They commented that bottle-feeding was "just easier." One teen, Ms. P, a seventeen-year-old mother of a three-month-old infant stated, "I just didn't want to breastfeed. I didn't have any breasts. They weren't large enough." Another teen, Ms. L, stated, "I didn't try breastfeeding because my milk never came in. I had inverted nipples and it (breast milk) never came out." Ms. V, a sixteen-year-old mother with a four-month-old infant, feared the pain and stated "I thought it was going to hurt more. My breasts got big

and they started to hurt. I just assumed it (breastfeeding) would hurt even more."

Teen mothers often verbalized a misunderstanding of proper breastfeeding technique. Unfortunately, after making only a few attempts, these teens tended to abandon their efforts and, in frustration, decided to discontinue breastfeeding. One teen, Ms. E, stated, "When I tried to breastfeed at home, he didn't want to be breastfed. He'd only suck for a little while" Another teen, Ms. H., stated, "It (breast milk) wasn't filling her up. The baby would cry and suck on my breast for more than 45 minutes." All of the teens surveyed (n=10), regardless of their EPDS score, reported that they were not currently breastfeeding their infant.

Describing Adolescents' Experiences with Birth Control Use

Nearly half of the teens (40%) surveyed disclosed that they were not using any form of birth control to prevent pregnancy. Teens' reasoning for not using birth control was varied and often unrealistic. Interestingly, the majority of the teens (n=4) who scored high on the EPDS revealed that they were currently using birth control to prevent pregnancy; whereas the majority of the teens (n=3) who scored less than nine on the EPDS, were found not to be currently using any form of birth control, thus indicating that feelings of depression did not influence their choice.

Three themes emerged when discussing the issue of birth control use among teen mothers: reproductive irresponsibility, an unwillingness to explore alternative methods of birth control, and a misunderstanding of birth control side effects. One teen, Ms. H, stated, "I'm not using any (birth control) because I'm not having sex." Another teen, Ms. E, stated, "I didn't get my shot (Depo-Provera) because I had to work. All I do is work. I don't have time for anything else. I don't think about it." Another example includes Ms. M, who stated, "I have a prescription for the pill but I haven't filled it."

Teens who had initially tried using birth control after pregnancy (n=2) were oftentimes unwilling to seek alternate methods of birth control if necessary. One teen, Ms. R, began to take the birth control pill but stopped because it made her sick. She was unfortunately unwilling to explore any alternative methods. When discussing birth control choices with another teen, Ms. E, she expressed disinterest in almost all of the alternatives discussed. After verbalizing her inability to find the time to make an appointment to receive her Depo-Provera injection, the researcher suggested oral contraceptives as a possible alternative. Ms. E, responded by stating, "I can't keep up with the pill." Teens also described side effects that were inaccurate when discussing their choices for birth control. This information revealed that many teens are misinformed about the side effects of various contraceptives. Ms. V, stated, "I didn't want to risk her (infant) falling out with the Depo shot. (fear of preterm labor???) My friend also told me that it causes hair loss."

An overview of all ten teens and their depression scores and decisions regarding breastfeeding and birth control use is displayed in Table 2.

Age	PPD	BC*	BF Initiate**	BF Current**
16	Yes	No	No	No
16	No	Yes	No	No
16	No	No	No	No
17	Yes	Yes	No	No
17	No	Yes	No	No
17	No	Yes	No	No
17	No	No	Yes	No
17	No	No	No	No
17	Yes	Yes	Yes	No
18	Yes	Yes	Yes	No
* BC = Birth control				
** BF = Breastfeeding				

Table 2. Teen Depression, Breastfeeding and Birth Control Results

DISCUSSION

The transition from adolescence to adulthood is a difficult journey that must be embarked upon by all teens at some point in their lives. The pregnant adolescent is often forced to make this arduous journey in a matter of months. The choices that a teen mother makes during this time affect both her life as well as that of her baby. Often the teen mother is depressed at postpartum due to various stressors, which intensify her feeling of being generally overwhelmed with a situation she is unprepared to handle. Without proper guidance and support, teen mothers may make erroneous decisions. While birth control use was not found to be influenced by depression, the choice of breastfeeding may be. Due to the small sample size, statistical correlations were not done; however, it is evident that educating adolescent mothers is paramount. Educational programs must be established to promote healthy maternal and infant outcomes among teen mothers. The prevention of poor maternal-infant bonding is of great importance in reducing the incidence of child neglect and abuse among teen mothers. Resources must be made available to those teen mothers who need assistance.

The HB program at the University of Texas at Arlington provides adolescent mothers with educational as well as emotional support. Teens in the HB program are offered one-on-one support by senior nursing students. Continued evaluation of the HB program is projected with future plans to include a comparison teen group without involvement in any community programs of a similar nature. Analysis of data that has been collected on the HB program indicates that teens needs include additional education regarding birth control use and breastfeeding instruction. Additional time for postpartum visitation and instruction is also recommended.

REFERENCES

Beck, C.T. (1992). The lived experience of postpartum depression: A phenomenological study. *Nursing Research, 41*(3), 166-170.

Beck, C.T. (1993). Teetering on the edge: A substantive theory of postpartum depression. *Nursing Research, 42*(1), 42-48.

Chen, C. (1996). Postpartum depression among adolescent mothers and adult mothers. *Kao Hsiung I Hsueh Ko Hsueh Tsa Chih, 12*(2), 104-113.

Chen, C., Chou, F., Tseng, Y., & Wang, S. (1999). Controlled study of postpartum depression in adolescence. *Nursing Research (China), 7*(5), 459-467.

Clifford, C., Day, A., Cox, J., & Werrett, J. (1999). A cross-cultural analysis of the use of the Edinburgh Post-Natal Depression Scale (EPDS) in health visiting practice. *Journal of Advanced Nursing, 30*(3), 655-664.

Mackey, M., & Tiller, C. (1998). Adolescents' description and management of pregnancy and preterm labor. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 27*(4), 410-419.

Morse, J.M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research, 40*(1), 120-122.

Nichols, F.H. & Zwelling, E. (1997). *Maternal-Newborn Nursing.* Philadelphia: W.B. Saunders Company.

Wambach, K.A. & Cole, C. (2000). Breastfeeding and adolescents. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 29*(3), 282-294.

Wisner, K. & Wheeler, S. (1994). Prevention of recurrent postpartum major depression. *Hospital & Community Psychiatry, 45*(12), 1191-1196.

Gender Discomfort Differences During Enema Administration

Allyson J. Schuerman

Mentor: Marilee Schmelzer PhD, RN, School of Nursing

ABSTRACT

Enemas are solutions administered rectally to stimulate defecation and are commonly used to treat severe consti-pation or clean the colon before diagnostic procedures and surgery. Although enemas are generally considered uncomfortable, very few studies have examined enema discomfort and possible gender differences have been ignored. The investigator performed a secondary analysis on data collected in a study comparing the effectiveness and side effects of three different solutions (soapsuds, tap water, polyethylene glycol electrolyte solution (PEG-ES)) (Schmelzer, Schiller, Chappell, Case, & Meyer, 2000). The purpose of this study was to describe discomfort experienced during enema administration and to determine if men and women experience the same type and amount of discomfort. It was expected that women would be more sensitive to enema administration than men. As a result of this study, no significant difference was evident between the discomforts the genders experienced, but other variations were found.

INTRODUCTION

Little is known about gender differences since clinical trials have traditionally in-cluded mostly men, due to fears of harming the fetus if a woman becomes pregnant (Miaskowski, 1997). Unfortunately when findings from studies of men have been applied to everyone in practice, it has become obvious that women and men are different. To rectify this situation, the National Institutes of Health (NIH) mandates that women be included in studies conducted by those who receive NIH grants (Miaskowski, 1997).

Preliminary studies suggest gender influences: (a) the type of pain symptoms ac-companying a particular disease (Everts, Bjorn, Wahrborg, Hedner & Herlitz, 1996; Milner, et al., 1999), (b) individual pain sensitivity and tolerance (Fillingim, Keefe, Light, Booker & Maixner, 1996; Fillingim, Edwards & Powell, 1999; Riley III, Robinson, Wise, Myers & Fillingim, 1997), and (c) the effectiveness of pain relief medications (Miaskowski, 1997; Roehr, 1998). Men and women may also respond differently to discomfort incurred by nursing procedures and the strategies used to reduce that discomfort.

Enemas are solutions administered rectally to stimulate defecation, and are com-monly used to treat severe constipation or clean the colon before diagnostic procedures and surgery. They are usually uncomfortable (Schmelzer & Wright, 1996). Very few studies have examined enema discomfort, and none have addressed gender differences. The pur-pose of this study is to describe enema discomfort and to determine if men and women experience a different type and intensity of discomfort when receiving large volume en-emas. By gaining this knowledge, nurses will be able to accurately assess enema discomfort and develop strategies to make enemas more comfortable.

BACKGROUND

Enema Discomfort

Although enemas are generally considered unpleasant, only four studies examine discomfort during enema administration. None addressed whether men and women experi-ence enema discomfort differently. Padilla and Baker (1972) asked subjects, "Were there any difficulties or discomforts with the enema(s)?" The researchers however did not report the results. Later, Duffin, Casteldon, and Chaudhry (1981) reported more detailed informa-

tion regarding the enema experience for patients. Half of the 50 subjects reported feeling better after an enema, two percent felt worse, and the others had no opinion. Participants also listed specific symptoms such as embarrassment (20 reports), pain (16 reports), and nausea (10 reports).

Enema technique may influence the amount and intensity of discomfort experienced by the recipient. Schmelzer & Wright (1996) asked 24 experienced registered nurses what they do to make enemas more comfortable. Most of the nurses reported individualizing comfort measures and said they regulate administration speed according to patient tolerance. All but one nurse agreed that the most comfortable enema is one that is warmed to body temperature and administered slowly. They thought the more comfortable enemas were easily retained and therefore more effective.

The contents of the enema solution administered could also influence the discomfort experienced. Researchers tested whether soapsuds caused more discomfort than tap water enemas in a sample of 25 patients scheduled for liver transplants (Schmelzer, Case, Chappell, & Wright 2000). After random assignment to two treatment groups, subjects received one liter of either soapsuds (13 subjects) or tap water (12 subjects) rectally. Data analysis showed no significant differences in the amount of discomfort. Collectively, the most common discomforts were abdominal cramping (8 reports), rectal fullness (6 reports), abdominal pain (2 reports), and embarrassment (2 reports) (Schmelzer, et al., 2000).

No one has determined if men and women experience differences in enema discomfort, but researchers have studied gender differences in other types of discomfort. The following sections will describe gender differences in experimentally induced pain, pain as a predictor of a physical problem, and pain during clinical procedures.

Experimentally Induced Pain

Researchers have studied the pain response by applying a painful stimulus and measuring pain threshold and tolerance. For example, heat was used as the pain stimulus in a study including 117 females and 92 males (Fillingim, et al., 1999). The heat was applied through a stimulator on the subjects' forearms. Subjects were instructed to press a button once when they first felt warmth, a second time when they felt pain (threshold), and a third time when the pain became intolerable (tolerance). The researchers found that females were more sensitive than males to the thermal stimuli (Fillingim, et al., 1996; Fillingim, et al., 1999). In addition, the researchers asked subjects to complete a questionnaire. Responses indicated that women visited health care providers more often and reported pain in more places on their bodies than the men did (Fillingim, et al., 1999).

Mechanical pressure has also been used to measure gender differences in pain tolerance and threshold. Researchers applied mechanical pressure to various body sites, for instance, fingers and achilles tendon, using motor driven rods and pressure stimulators and found women were more sensitive to mechanical stimuli than men (Riley III et al., 1998). Since enemas can be given at different temperatures, and the large volume exerts mechanical pressure on the walls of the colon, it seems reasonable that gender differences in enema sensation would exist.

Pain as a Symptom of a Disease

Unfortunately, no one is immune from pain, and when it arises it may indicate a medical problem. An acute myocardial infarction (AMI) is a life threatening condition that

demands immediate hospitalization. The most recognized symptom of an AMI is severe chest pain, but women may have more atypical symptoms than men (Everts, et al., 1996). In a study of 455 patients with a confirmed AMI, Everts, et al., (1996) found only one difference between the genders: women reported more neck and back pain than men did.

In a second study, researchers studied 217 subjects (90 women, 127 men) with coronary heart disease (CHD) about their symptoms of an AMI. The findings revealed significant gender differences among the non-chest symptoms but not the chest symptoms. Women reported nausea, vomiting, dyspnea, palpitation, indigestion, and midback pain more frequently than men (Milner, et al., 1999). This difference in the pain experienced by women and men may explain why women with AMI are often overlooked and less frequently admitted to coronary care units (Everts, et al., (1996; Milner, et al., 1999). Obviously, the importance of recognizing common and unusual symptoms of pain is critical to the diagnosis and treatment of the physical problem.

In a study of 205 in-hospital patients, researchers asked participants to respond to a questionnaire regarding their pain. As a result, females reported pain as "distressing to excruciating" significantly more frequently than males (Yates, Dewar, Edwards, Fentiman, Najman, Nash, Richardson & Fraser, 1998). Men and women with the same pathology were compared, and the researchers concluded women reported greater pain than men (Miaskowski, 1997). According to Vallerand (1995), men experience more backaches, cardiac pain, and cluster headaches while women suffer more from migraine headaches, arthritis and facial pain. If pain symptoms of a disease are different, a discomfort difference during enema administration may exist.

Pain During a Clinical Procedure

Men and women may experience pain differently during an invasive procedure. A research team studied a group of 48 men (21) and women (27) without teeth undergoing oral surgery, which consisted of implanting a small titanium rod in their gums (Morin, Lund, Villarroel, Clikie & Feine, 1999). The subjects were required to keep a diary and to rate the pain intensity they experienced immediately after the surgery. The participants reported this information three times a day for 14 days after the surgery. During the postoperative phase, researchers found no difference between the pain men and women experienced. The only difference discovered was that women used more expressive words to describe their pain (Morin, et al., 2000).

In addition, researchers asked 110 children ages 3-15 years of age to report their unpleasantness and pain immediately after a venipuncture. No gender differences occurred among the younger children, but at age 8 and older a gender difference emerged. Older girls reported more unpleasantness and a greater pain intensity by the venipuncture than the boys of the same age (Goodenough, et al., 1998). Similarly in an immunology clinic 62 children ages 3-12 years old were asked to rate anticipatory pain before allergy testing. The results showed girls reported more pain than boys. Girls also expressed that positive medical experiences involved less pain (Carr, Lemanek, & Armstrong, 1998). Since studies demonstrated a gender difference expressing pain in the clinical setting due to a procedure, a difference may be evident during enema administration.

Hormonal Influence

Although there is only preliminary data, it is reasonable that gender differences will exist when one considers the effects of hormonal differences on other body systems. For

example, female and male central nervous systems respond differently to estrogen and testosterone (Roehr, 1998), and researchers found that a kappa-opoid analgesic provided significantly greater analgesia in females (Miaskowski, 1997). During the different phases of a woman's menstrual cycle, neurotransmitters have been found to vary (Blum, et al., 1992; Anderberg, U., 2000). It is possible that women may experience a completely different pain perception to the same stimuli depending on the phase of her menstrual cycle.

Other Influential Factors

Age, previous pain, and socio-cultural factors may also influence the pain response. While babies demonstrate an intense reaction to pain, older people react less intensely because they have developed effective coping strategies. Also, women who have given birth to a child may find it easier to cope with pain than those who have not (Carr, 1997). Many believe it is more socially acceptable for a woman to express pain. A man who expresses pain may be considered weak (Carr, 1997). Culture shapes people and may determine the actions and responses to stimuli. For instance, Carr (1997) mentions Martin & Belcher's (1986) finding that American nurses interpreted screaming as severe pain, while South African nurses thought quietness was a sign of even greater pain.

METHODOLOGY

This study was a secondary analysis of data collected by Schmelzer, Schiller, et al., (2000). In the original study the researchers gave a group of 24 healthy subjects a different enema once a week for 3 weeks. The three solutions were soapsuds, tap water, and polyethylene glycol electrolyte solution (PEG-ES). Schmelzer, Schiller, et al., (2000) compared the safety and effectiveness of the three solutions. One of the variables measured during the enema procedure was discomfort.

The purpose of this secondary data analysis was to examine the enema discomfort data to determine if gender differences existed. It was hypothesized that female subjects would experience significantly greater discomfort during enema administration than the males ($\alpha < 0.05$).

Setting and Sample

The original study was conducted in a gastrointestinal laboratory at a large medical center in north central Texas. The facility included private rooms with easy access to a private restroom.

A convenience sample of 24 was selected to ensure that there were equal numbers of men and women, with a racial background of four Hispanics, three African Americans and seventeen Caucasians. The participants had to be at least eighteen years of age, mentally alert, and healthy. The principal investigator (PI) screened potential subjects to exclude those with bleeding disorders, diabetes, ileostomies, colostomies, inflammatory bowel disease, heart disease, renal disease, or any gastrointestinal abnormalities. She instructed the subjects not to consume ibuprofen or aspirin products one week prior to and throughout the study (due to possible bleeding from biopsies), and to maintain their usual diet throughout the three weeks of the study.

Intervention

To eliminate the effects of order, the PI randomly assigned subjects to one of six groups with four participants in each group. Each group received all three types of enemas;

the only difference was the order in which they were given. The enema solutions contained 1100 grams of either soapsuds (6 grams of castile soap per liter of deionized water), tap water (deionized water), or polyethylene glycol electrolyte solution (PEG-ES).

Data collection

Schmelzer, Schiller, et al., (2000), created a visual analogue scale to rate the different discomforts of an enema (see Appendix 5). The visual analogue scale consisted of 11 lines with a place for additional comments at the bottom. Each line was ten centimeters long and was labeled from "no discomfort" on the far left to "severe discomfort" on the far right. When describing their discomfort, the subjects inserted a slash mark at the point on the line which best corresponded to the amount of discomfort they experienced for each symptom. The research assistant measured the marks on the line to the nearest half-centimeter and recorded the results. Subjects were also asked to describe any additional discomfort at the bottom of the scale.

Validity was supported by previous studies (Duffin, et al., 1981, Schmelzer & Wright, 1996) and reliability was tested in the Schmelzer, Schiller, et al., study (2000). The researcher used nine of the eleven types of discomfort measured in the original study in the secondary data analysis (embarrassment, nausea, vomiting, abdominal cramps, abdominal pain, rectal itching, rectal, burning, dizziness and feeling faint).

Researchers recorded demographic data (gender, age, race), blood pressure, medication use, bowel habits and the amount of enema administered.

Procedure

The PI recruited subjects by word of mouth at the medical center and at a local university and determined if they met study criteria. Subjects participated in the study for three Saturdays in a row. The first week, subjects gave informed consent, and then emptied their bladders. After taking vital signs, a research nurse escorted subjects to the treatment room where she gave the enema according to a standard procedure. One registered nurse (female) gave all the enemas and a second registered nurse (female) prepared and warmed the enema solutions, and measured retention time. The subjects chose music to play to help them relax. Researchers instructed the subjects to hold the enema as long as possible but not more than ten minutes, and to eliminate the solution in an adjacent bathroom. Immediately after eliminating the enema, subjects rated the intensity of their discomfort using a visual analogue scale for discomfort (Appendix 5). The subjects were paid for participating in the study.

FINDINGS

Description of the sample

Twenty-six people were recruited and two subjects dropped out leaving a final sample of 24. The majority of the sample was young. Twenty were younger than 37 but three women were in their forties and one woman was 65. All the participants reported that they were in good health, but some were taking medications daily. Five of the twelve men and all the women reported using at least one prescription medication (Appendix 1). When asked about usual bowel habits, 18 reported at least one bowel movement daily and six reported at least one every other day. Nine subjects had experienced at least one enema in the past (Appendix 1). The men and women had no significant differences in bowel elimination patterns or previous enema use.

Significant differences were found between the blood pressures of men and women taken immediately before the enema was given. The men's mean systolic blood pressure of 127 mm Hg was significantly higher than the women's mean systolic pressure of 112 mm Hg (t:4.97, df: 22, p=.000). The men's mean diastolic pressure of 88 mm Hg was significantly higher than the women's diastolic blood pressure of 79 mm Hg (t:2.82, df: 22, p=0.01). Interestingly, some studies have shown that there is diminished pain sensitivity in hypertensive people (Nyklicek, Vingerhoets & Van Heck, 1998).

All but two people were able to take all enemas entirely. One woman took only 381 ml of tap water and 654 ml of the PEG-ES, although she did take the entire soapsuds enema. One man took only 974 ml of PEG-ES, although he took all of the other two enemas.

Hypothesis Testing

Gender comparisons of the nine types of discomfort for each type of solution are found in Appendix 2. The most common discomforts were abdominal cramps, abdominal pain, embarrassment and nausea. The ratings for each type of discomfort were added to obtain a total discomfort score for each individual. The mean total discomfort scores are found at the bottom of Appendix 2.

It was hypothesized that female subjects would report higher pain intensity than the males during enema administration of the three different solutions. The T-test was used to analyze the data and compare the overall discomfort between males and females. No significant difference was found between the genders at the 0.05 level of significance (Appendix 3). The difference in discomfort between enema solutions will be addressed by Schmelzer, Schiller, et al., (2000).

Although no gender differences were found in the total discomfort scores, women tended to report more types of discomfort (Appendix 4). The men experienced no vomiting or rectal itching with the soapsuds solution; no vomiting, dizziness, or faintness from the tap water solution; and no vomiting, rectal burning, dizziness or faintness from the PEG-ES solution (Appendix 2). At least one woman experienced each of these discomforts (Appendix 4).

Qualitative analysis of the subjective comments found that the women listed a greater variety of symptoms and described all symptoms more vividly. For example, four women commented about the enemas' effects on urination and three said they had trouble holding the enema.

Subjects also identified symptoms not listed on the visual analogue scale, including: the inability to hold in the solution, tingling of the abdomen, diaphoresis (sweating), pressure extending to the throat and a "stomach washing" sensation.

DISCUSSION

The study found no significant gender differences, but qualitative data indicated that the women tended to describe their discomfort more thoroughly than the men. The researcher identified four limitations. First, it was assumed that people would be honest, but, in fact, they might have been unwilling to admit the actual degree of discomfort. Although this may be a problem, pain measurement is generally considered to be "what the patient says it is" (McCaffery & Pasero, 1999).

Second, a female gave all the enemas. Previous studies have found that willingness to describe or admit to pain is influenced by gender of the experimenter (Miaskowski, 1997).

Third, the researchers deliberately made the enema as comfortable as possible and adjusted the administration speed according to the subjects' feedback. Administering the enemas at different speeds may have introduced variability into the study, and reduced the ability to find a difference between groups.

Fourth, the original study was not designed to investigate gender differences. Its repeated measures design provided adequate power even with a relatively small sample size. However, the power of a repeated measures design was lost when gender differences were examined. Thus, the sample may have been too small to detect a difference.

Recommendations for Further Study

Several areas need further research. For example, procedural pain, gender differences and enemas all require more thorough investigation. Gender differences in enema discomfort may become evident if a larger sample is used. Moreover, acute pain and discomfort during other nursing procedures should be investigated.

During data analysis, the researcher discovered problems with the visual analogue scale. The scale did not differentiate between discomforts experienced while the solution was infusing into the rectum, discomfort experienced during the ten minutes when it was held and discomfort during and after defecation. The study also identified other symptoms to add to the tool.

In addition, the significantly higher blood pressures exhibited by the male participants may be due to increased levels of anxiety (Craven & Hirnle, 1996). It may be valuable to have participants rate their level of anxiety prior to enema administration in future studies.

In conclusion, during enema administration, no significant discomfort differences became evident between the genders. Based on qualitative data, women did express additional comments about discomforts more vividly and more often than men and used more prescription medication. Finally, further investigation is required and no generalizations can be made at this time.

ACKNOWLEDGEMENTS

REFERENCES

Anderberg, U. M. (2000). Upsala, Sweden [Letter to the editor]. Pain. 87 (2000), 109-111. Retrieved June 14, 2000 from World Wide Web: http://www.elsevier.nl/locate/pain

Carr, E. (1997). Factors influencing the experience of pain. Nursing Times, 93 (39), 53-54.

Carr, T. D., Lemanek, K. L., Armstrong, F. D. (1998). Pain and fear ratings: clinical implications of age and gender differences. Journal of Pain and Symptom Management 15 (5) 305-313.

Craven, R. F., Hirnle, C. J. Fundamental of Nursing: Human Health and Function (2nd ed.). Philadelphia: Lippincott.

Duffin, H. M., Castleden, C. M., & Chaudhry, A. Y. (1981). Are enemas necessary? Nursing Times, 77 (45), 1940-1941.

Everts, B., Karlson, B., Wahrborg, P., Hedner, T. & Herlitz, J. (1996) Localization of pain in suspected acute myocardial infarction in relation to final diagnosis, age and sex and site and type of infarction. Heart & Lung Journal of Acute & Critical Care 25 (6) 430-437.

Fillingim, R., Keefe, F., Light, K., Booker, D. & Maixner, W. (1996) The influence of gender and psychological factors on pain perception. Journal of Gender, Culture and Health 1 (1) 21-36.

Goodenough, B., Thomas, W., Champion, G. D., Perrott, D., Taplin, J. E., von Baeyer, C. L., Ziegler, J B. (1999). Unraveling age effects and sex differences in needle pain: ratings of sensory intensity and unpleasantness of venipunture by children and their parents. Pain (80) 179-190.

McCaffery, M., & Ferrell, B., (1992). Does the gender gap affect your pain-control decisions? Nursing, 92, August, 48-51.

McCaffery, M., Pasero, C. (1999). Pain: Clinical Manual (2nd ed.). St. Louis: Mosby.

Miaskowski, C. (1997). Women and pain. Critical Care Nursing Clinics of North America, 9(4), 453-457.

Morin, C., Lund, J. P., Villarroel, T., Clokie, C.M.L., & Feine, J. S. (1999, September). Differences between the sexes in post-surgical pain. Pain. 85 (2000), 79-85. Retrieved June 14,2000 from World Wide Web: http://www.elsevier.nl/locate/pain

Nyklicek, I., Vingerhoets, A. J. J. M., Van Heck, G. L. (1998). Hypertension and pain sensitivity: effects of gender and cardiovascular reactivity. Biological Psychology, (50), 127-142.

Padilla, G., & Baker, V. (1972). Variables affecting the preparation of the bowel for radiologic examination. Nursing Research, 21 (4), 305-312.

Roehr, B. (1998). The puzzle of gender and pain. Journal of the International Association of Physicians in AIDS Care, 4 (6), 23-27.

Schmelzer, M., Case, P., Chappell, S. M., & Wright, K., B. (2000). Colonic cleansing, fluid absorption, and discomfort following tap water and soapsuds enemas. Applied Nursing Research, 13 (1), 1-10

Schmelzer, M., Schiller, L., Chappell, S., Case, P. & Meyer, R. (2000).[Safety and effectiveness of large volume enema solutions]. Unpublished raw data.

Schmelzer, M., Wright, K., (1996). Enema administration techniques used by experienced registered nurses. Gastroenterology Nursing, 19 (5), 172-175.

Vallerand, A. H., (1995). Gender differences in pain. Image - the Journal of Nursing Scholarship 27(3), 235-237.

Yates, Dewar, Edwards, Fentiman Nash Fraser et al., (1998). The prevalence and perception of pain amongst hospital in-patients. Journal of Clinical Nursing 7 (6) 521-530.

APPENDICES

Appendix 1

Variables	Males		Females
Age			
M	26.5		34.3
Range	20-34		20-65
Total Reports of Prescription Use			
Inhalers	2	Birth Control	2
Antihistamine	1	Hormones	2
Antibiotics	1	Antibiotics/Antiinfective	3
Other	0	Other	5
Total Reports of Over-the-Counter Use			
Tylenol	1	Tylenol	2
Vitamins	4	Vitamins	5
Bowel Elimination Frequency			
One / daily	5		6
Two / daily	2		2
Three / daily	2		1
Once every other day	2		0
Once every 1-2 days	1		3
Previous Enema Use			
Yes		4	5
No		8	7

Description of Several Reported Variables for Men and Women

Discomfort		Soapsuds		Tap Water		PEG-ES	
		Males	Females	Males	Females	Males	Females
Embarrassment	Mean	0.9	1.6	1.0	1.5	0.5	1.9
	SD	1.7	2.6	1.4	2.2	1.1	3.0
	Range	0-5.5	0-8.5	0-4.0	0-7.5	0-4.0	0-9.
Nausea	Mean	0.6	1.7	0.3	0.8	0.1	0.3
	SD	1.4	2.0	0.4	2.1	0.2	0.6
	Range	0-5.0	0-5.5	0-1.0	0-7.5	0-0.5	0-2.0
Vomiting	Mean	0	0.1	0	0.2	0	0.1
	SD	0	0.2	0	0.2	0.1	0.2
	Range	0	0-0.5	0	0-0.5	0-0.5	0-0.5
Abdominal Cramps	Mean	4.6	4.6	4.0	2.5	1.6	1.4
	SD	3.4	2.9	2.7	2.7	1.6	1.7
	Range	0-8.0	.5-9.5	0-7.5	0-7.5	0-5.0	0-5.0
Abdominal Pain	Mean	2.8	2.7	2.9	1.9	0.8	0.4
	SD	3.2	3.2	2.9	2.1	1.7	0.4
	Range	0-8.0	0-9.5	0-7.5	0-6.0	0-5.0	0-1.0
Rectal Itching	Mean	0	0.1	0.1	0.1	0.1	0.1
	SD	0	0.2	0.3	0.2	0.3	0.2
	Range	0	0-0.5	0-1.0	0-0.5	0-1.0	0-0.5
Rectal Burning	Mean	0.2	0.4	0.1	0.3	0	0.2
	SD	0.5	0.9	0.3	0.6	0	0.2
	Range	0-1.5	0-2.5	0-1.0	0-2.0	0	0-0.5
Dizziness	Mean	0.5	0.4	0	0.6	0	0.1
	SD	1.9	0.5	0.1	1.3	0	0.2
	Range	0-6.5	0-1.5	0-0.5	0-4.5	0	0-0.5
Faintness	Mean	0.5	0.3	0	0.3	0	0.2
	SD	1.6	0.5	0	0.3	0	0.3
	Range	0-5.5	0-1.5	0	0-1.0	0	0-1.0
Total Discomfort	Mean	10.1	1.9	8.5	8.0	3.2	4.7
	SD	8.0	8.3	5.4	7.8	3.2	4.8
	Range	0-23.5	0.5-27.5	0-15.0	0.5-28.0	0-11.0	0-15.5

**Gender Comparison of the Discomfort Scale Scores
(According to Solution Given)**

	Gender	M	SD	t	df	Sig (2tail)
Soapsuds Discomfort	M	10.1	8.0	-.536	22	.6
	F	11.9	8.3			
Tap Water Discomfort	M	8.5	5.4	.183	22	.9
	F	8.0	7.8			
PEG-ES Discomfort	M	3.2	3.2	.896	22	.4
	F	4.7	4.8			

Total Discomfort Comparison Between Males and Females

Appendix 4

TYPES OF DISCOMFORT	Type of Solution					
	Soapsuds		Tap Water		PEG-ES	
	Males	Females	Males	Females	Males	Females
Embarrassment	5	7	6	9	3	6
Nausea	4	7	4	4	2	5
Vomiting	0	3	0	4	1	3
Abdominal Cramps	11	12	11	9	10	7
Abdominal Pain	8	8	9	9	4	6
Rectal Itching	0	3	2	3	2	2
Rectal Burning	2	2	2	5	0	4
Dizziness	1	5	1	5	0	3
Faintness	1	5	0	6	0	3
TOTAL DISCOMFORT	32	52	35	54	22	39

Frequency of Men and Women Reporting at Least Some Degree of Discomfort

Appendix 5

Subject #_____   Time_____

Please place a mark through the line to indicate the intensity of your discomfort during the enema?

No embarrassment _____Severe embarrassment

No nausea _____Severe nausea

No vomiting _____Severe vomiting

No abdominal cramps_____Severe abdominal cramps

No abdominal pain _____Severe abdominal pain

No rectal itching _____Severe rectal itching

No rectal burning _____Severe rectal burning

No dizziness _____Severe dizziness

No feeling of faintness _____Severe feeling of faintness

No soiling from stool leakage_____Great soiling from stool leakage

No leakage of urine_____Much leakage of urine

Did you experience any additional discomforts? If so, please describe them.

Subject Reports of Discomfort from Enemas

Thermodynamic Studies of the P4-P6 Domain of the Group One Intron

Tina Holder

Mentor: Dr. Thomas Gluick, Department of Chemistry and Biochemistry

ABSTRACT

An attempt was made to calculate the change in Gibbs free energy (DG) for the secondary to tertiary folding in the group I ribozyme domain, P4-P6. This was first attempted using UV spectroscopy melting curves in the hope that mutants lacking the tertiary structure would have very different melting profiles than the wild type P4-P6 domain. Studies of the profiles showed cooperativity in the secondary and tertiary folding which prevents direct calculation of free energy changes from the melting profile. Buffer conditions under which the melts were being performed were optimized to ensure that secondary structure was stabilized only by K+ ions and that tertiary structure was stabilized by specific Mg+2 interactions. The next step was to take the optimized buffer conditions and perform calorimetry experiments that gave direct experimental values for the DH from which the free energy change can be calculated. At this point the enthalpy values found were too high when compared to theoretically predicted values. Possible causes include aggregation or hydrolysis of the RNA molecules.

INTRODUCTION

RNA molecules have come under greater study in the past twenty years, as their role in cellular function and their ability to catalyze reactions have been explored more deeply. RNA molecules perform a variety of intracellular functions that are essential to cell survival. Messenger RNA molecules (mRNA) are made as mimics of the DNA sequence and then are translated into proteins by the ribosome. Transfer RNAs (tRNA) are responsible for delivering the amino acids that are being added to the peptide chain into the proper position. Ribosomal RNA (rRNA), along with proteins, actually makes up the ribosome, which builds new proteins. RNA molecules are also found in other forms in the cell and contribute to many other activities.

Due to their essential and ubiquitous presence in living organisms, RNA molecules inherently have interesting chemical properties. This can be seen in the wide variety of roles that RNA molecules can fill and in the varied interactions that they can have with proteins. To truly appreciate the biomolecular abilities of RNA, the chemical basis for that ability must be understood. One way to study the molecule is to describe it in thermodynamic terms and to find values for the thermodynamic transitions that the molecule is undergoing. The value of the change in enthalpy (ΔH), the change in Gibbs' free energy (ΔG), and the change in entropy (ΔS) all give important information about the probability and direction in which a reaction will go. The enthalpy change is representative of the amount of heat the molecule absorbs or releases as it undergoes a transition or reaction. Entropy is the amount of increase or decrease of disorder in the system under study. The Gibb's free energy is found by taking a combination of entropy and enthalpy according to Equation 1. It is the value that dictates if the reaction or transition will happen spontaneously.

Equation 1: $\quad \Delta G = \Delta H - T\,\Delta S$

The work described centers upon a small domain of a larger mRNA that has catalytic activity. The mRNA is called the Group One Intron and was isolated from a unicellular organism Tetrahymena thermophila. This was one of the first RNA molecules found to have catalytic ability. Before this time, it was assumed that the only catalytically active molecules in cells were proteins. The discovery of catalytic RNA molecules, termed ribozymes,

gave new insight into the origin of biomolecular catalysis. The Group One Intron has been extensively studied to understand better how RNA molecules could perform catalytic activities when their relatively simple structure was compared to more complex protein structures. It was found that secondary and tertiary folding occurred that allowed for an active site to be formed. Our studies take the P4-P6 domain of this Group One Intron and, using calorimetry and UV spectroscopy, try to decipher the thermodynamic energies involving the secondary to tertiary transitions in the molecule. The Group One Intron is too complex to study in this manner, so we hope to gain a better understanding of its behavior by studying a smaller portion of the molecule.

To understand the purpose of the experiments one must understand the various levels of folding seen in RNA molecules. The actual sequence of the molecule is the primary structure which can loosely be pictured as a strand of bases with negative charges on its backbone and hydrophobic, aromatic bases attached to the backbone. In the environment of the cell, the RNA molecule is more stable when it folds upon itself and interactions form among the chain. Secondary structure occurs when specific hydrogen bonding and base stacking interactions form among the aromatic bases. The hydrogen bonding is fairly standard and thus makes predicting secondary structure fairly straightforward. Tertiary structure occurs when the unbound loops and bulges interact to form complex semiglobular folding. Tertiary interactions are much harder to predict, and by understanding the thermodynamic values for these transitions in well known RNA molecules, predictions can be made about less studied RNAs.

METHODS AND MATERIALS

RNA sample preparation

RNA molecules were transcribed using the Gurevich protocols off of DNA plasmids supplied by Dr. Sarah Woodson of Johns Hopkins University. There were five plasmids provided: one coded for the wild type P4-P6 and the other four mutants in which sites essential for tertiary structure had been mutated. One point of the work was to show that, indeed, the mutants had no tertiary structure. The RNA was concentrated on centricon spin columns and then ran on 8% urea acrylamide gels. The bands were cut out with a razor blade and then placed in an electroelution apparatus. A voltage was applied across the RNA bands and, since RNA is negatively charged, the molecules eluted into a receiving chamber at the positive end of the apparatus. From here the RNA was removed and placed in a corex tube to which 10 ml of ethanol and 300 ul of 10 M ammonium acetate were added. The tubes were placed at -20°C for twenty-four hours and then centrifuged at 8000 g for 20 minutes to precipitate the RNA molecules. The supernatant was disposed of and the pellets were rinsed with a 70% ethanol solution for five minutes at 8000 g. The supernatant was again removed and the pellets were dried under vacuum. After drying the pellets were resuspended in a MOPS buffer at pH 7 and used in UV melts. For calorimetry studies the RNA was dialyzed into the desired buffers.

UV Spectroscopy Studies

The RNA was added to the desired buffer to give a UV spectral optical density of 0.4 O.D. This usually meant adding one to ten microliters of the RNA solution to 1 to 1.4 ml of the buffer and mixing the solution well. The RNA was heated for ten minutes at 65°C to melt all of the secondary and tertiary structure and remove irregularities due to several

states of folding in the molecule. The typical melt consisted of a cooling phase in which the RNA was taken from 65°C to 15°C at the rate of 0.25°C to 1°C per minute. This phase allowed for the RNA molecules to reassume secondary and tertiary structure. The molecules were then heated from 15°C to 95°C at the same heating rate. Melting curves similar to those in Figure 1 show the increasing rate of UV light absorbance as the RNA molecule unfolds exposing the aromatic bases.

By monitoring this change in absorbance, we can learn about the structure of the RNA at various temperatures and can make some approximations of the transition energies. The transition energies found using UV spectroscopy are made by assuming several factors about the behavior of the molecule during the transition. The first is that the DH is independent of temperature change as the sample is heated. The second is in assumptions concerning the molecularity of the reaction. For the study of P4-P6, it was considered better to optimize the initial buffer conditions using UV spectroscopy and then find direct values for the DH through calorimetric experiments (which do not make the same assumptions as the spectroscopic measurements do). From the calorimetric value for the change in enthalpy the changes in entropy and Gibb's free energy can be calculated.

The interest in the P4-P6 RNA is in its transition from tertiary to secondary structure. The initial melts relied upon altering salt concentrations to optimize the secondary and tertiary study. Monovalent ions with a +1 charge (like K+1) nonspecifically stabilize the negative backbone of the RNA molecule. Divalent ions with a +2 charge (like Mg+2) may specifically stabilize tertiary structure. First, potassium concentrations were altered to fully stabilize the secondary structure without contributing to tertiary structure stability. Next titrations with magnesium were done to establish the best concentrations for tertiary structure stability. At this point by adding or taking magnesium away, the tertiary structure could be added or removed. In Figure 1 the melting curve of the wild type P4-P6 RNA molecule can be seen in the presence and absence of magnesium. From the melting curves slight differences can be seen in the slope, but the best way to display the data is to take the derivative of the curve giving a plot known as the melting profile. Figure 2 shows the melting profiles of the P4-P6 with and without magnesium. Distinct differences in height and width can be seen. These differences showed that in the sample with no magnesium little or no tertiary structure existed.
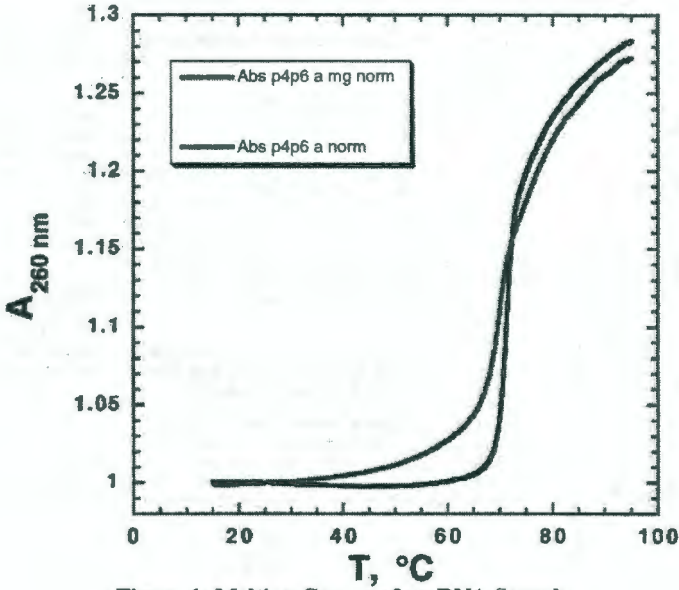
Figure 1: Melting Curves of an RNA Sample
Comparison of curves shows Mg^{+2} stabilizes RNA
with Mg^{+2} (in black and curves ≈ 1.01) and
without Mg^{+2} (in gray and curves ≈ 1.03)
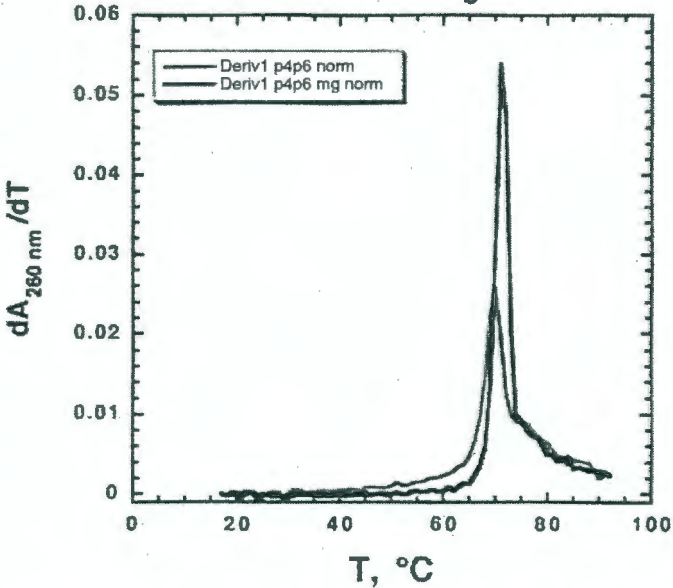
Figure 2: Melting Profiles of the Wild Type with and without Mg^{+2}
Melting profiles show cooperativity in unfolding increases with Mg^{+2}
Profile with Mg^{+2} (in gray and spikes $\approx .026$) and
without Mg^{+2} (in black and spikes $\approx .054$)

The next step was to compare the mutants that have been altered to have no tertiary structure using the same buffer conditions as above. In the presence or absence of magnesium, no tertiary structure should be present since the base interactions needed have been disrupted by mutation. Figure 3 shows melting profiles of one of the mutants compared with that of the wild type molecule in the presence of Mg+2. It can be seen that the mutant profile is much shorter, thus confirming that tertiary structure is absent in the mutated RNA molecule.
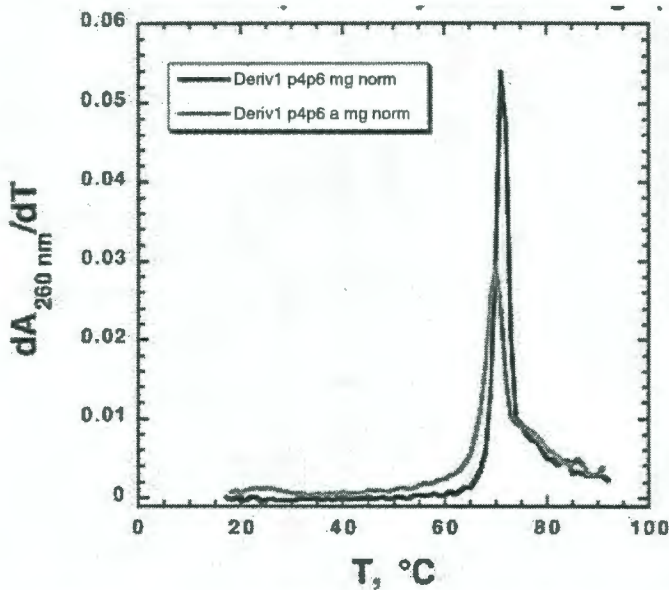
Figure 3: Melting Profiles of the Wild Type and Mutant in the Presence of Mg^{+2}
Melting profiles of RNA with tertiary structure in Mg^{+2} (in black and spikes ≈ .054) and RNA with disrupted tertiary structure Mg^{+2} (in gray and spikes ≈ .030)

Calorimetry Studies

At this time with optimized buffer conditions, calorimetry was started. The hope was that, as seen by Equations 2-4, we could calculate the change in enthalpy from the difference between the experimental values of heat absorbed by the RNA upon unfolding. The RNAs with no tertiary structure should be very similar in value and the RNA with tertiary structure should be greater to account for the extra energy needed to unfold the tertiary structure.

Equation 2: $\Delta H \text{ total} = \Delta H2 + \Delta H3$
(for the wild type P4-P6 with Mg^{+2})

Equation 3: $\Delta H \text{ total} = \Delta H2$
(for the wild type P4-P6 with no Mg^{+2})

Equation 4: $\Delta H \text{ total} = \Delta H2$
(for the mutant RNAs with no tertiary structure with or without Mg^{+2})

The buffer chosen to use for calorimetry was a 1M KCl solution with 10 mM potassium cacodylate used as a buffering agent. For the studies of Mg^{+2} containing samples the buffer was made to contain 20 mM $MgCl2$ in addition to the other salts. The samples were at greater than a milligram per milliliter concentrations. Samples were dialyzed into the desired buffer and degassed by stirring under vacuum before calorimetry. The calorimeter used was a Microcal calorimeter.

Differential scanning calorimetry is based on the comparison of a sample cell to a reference cell. The reference buffer is injected into one compartment and a sample under study is injected into a neighboring compartment. The samples are then heated, keeping both at the same temperature, from 10°C to 120°C. During the heating, the amount of electrical energy sent to the samples is monitored. The compartment containing the RNA sample requires more electrical energy than the reference buffer to remain at the same temperature because, as the molecule unfolds, it absorbs heat. The difference in electrical energy can be converted to the heat capacity of the molecule. Since, as seen in Figure 4, the DH is equal to the area under the curve, its direct value can be found.
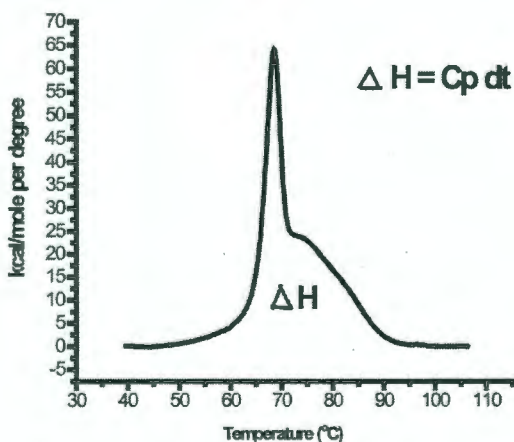
Figure 4: An Example of a Calorimetry Curve

Figure 4 shows calorimetry data obtained for the wild type P4-P6 RNA in the presence and absence of magnesium. As with the spectroscopy data, differences can be seen in the height and shape of the peaks. When integrated, values of 724 Kcal/mol for the magnesium containing sample and 692 Kcal/mol for the magnesium-free sample were found. At first, this seemed promising in that calorimetric data for the two differed and that the sample with magnesium, and thus tertiary structure, was larger.
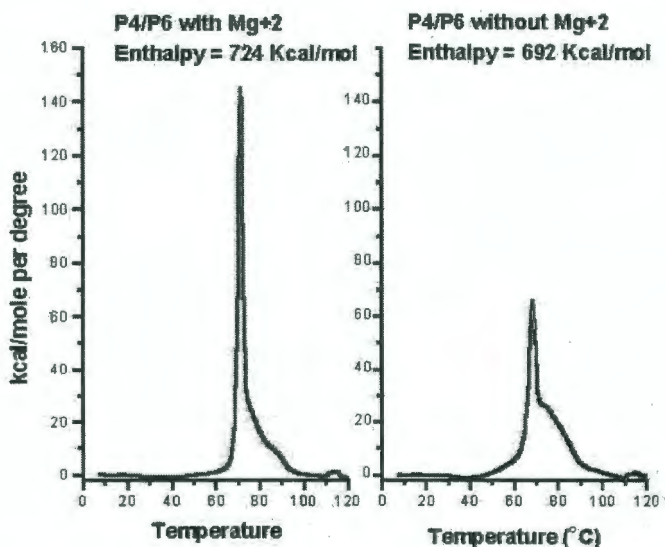
Figure 5: A Comparison of the P4-P6 RNA with and without Mg+2

The data showed that more energy was required to unfold the molecule in the presence of magnesium due to tertiary structure and that potentially the difference could be the DH of the tertiary transition. However before any experimental work was done, theoretical values for the DH were calculated. The estimated transition energy is predicted to be around 350-400 Kcal/mol based on a set of rules developed by Dr. Douglas Turner for estimating RNA thermodynamic energies. The fact that the experimental values were found to be about twice the theoretical values implied contributions to the energy due to other factors besides unfolding of the RNA. Possible causes include aggregation or hydrolysis. The RNA samples are at very high concentrations, and thus part of the energy seen may be due to intermolecular interactions rather than intramolecular interactions. Thus the initial energy is expended in breaking up the aggregated RNA molecules before actual melting can occur. Another factor may be hydrolysis or degradation of the RNA sample due to the high temperatures used in the calorimetry experiments.

Figure 6 shows calorimetry work performed on the mutant RNA samples. It was expected that the samples in the presence or absence of magnesium would have similar transition values, as no tertiary structure is present. The values were expected to be close to the wild type sample without magnesium as it, too, would have little or no tertiary structure. However, as can be seen from the calorimetry data, the shape of the mutant curves differed, and the enthalpy values were different from each other and from the wild type P4-P6 in the absence of magnesium.
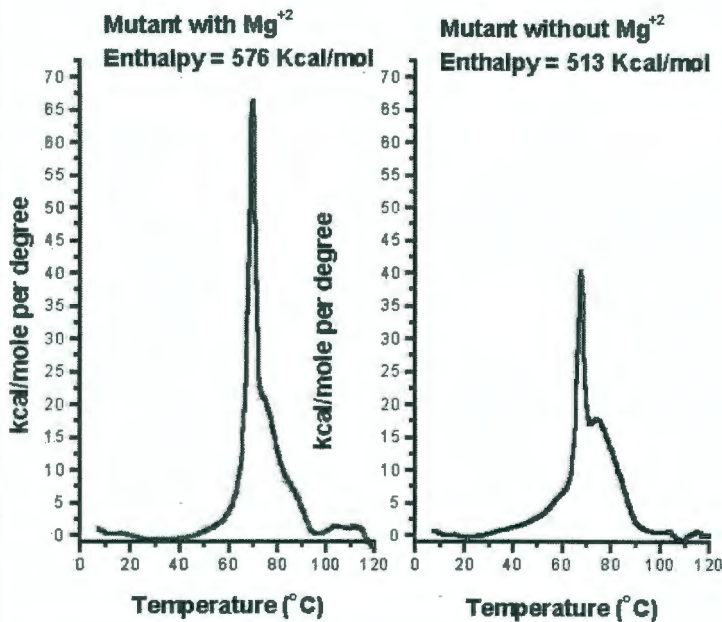
Figure 6: Calorimetry Scans of the Mutant RNA Molecules with and without Mg^{+2}

Conclusion and Future Work

At this time, work centers on resolving the cause of the extra energy seen in the calorimetry data. Steps include lowering the concentration of the RNA to reduce aggregation of the sample and lowering the salt concentrations to allow for the calorimetry to be conducted at lower temperatures with less hydrolysis. The numerical data obtained is meaningless until the large value of all measurements can be explained. Other work will include running more spectroscopy experiments to test other salt conditions.

REFERENCES

1. Chastain, Micheal, and Ignacio Tinoco, Jr. "Structural Elements in RNA." Progress in Nucleic Acid Research and Molecular Biology 41 (1991): 131-175.

2. Turner, Douglas H., and Martin J. Serra. "Predicting Thermodynamic Properties of RNA." Methods in Enzymology 259 (1995): 242-261.

3. Cate, Jamie H., et al. "Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing." Science 273 (1999): 52-82.

Morphine Microinjection into the Anterior Cingulate Cortex and the Decrease of Inflammatory Pain in the Rat

Damien Kinzler

Mentor: Dr. Perry N. Fuchs, Department of Psychology

ABSTRACT

The anterior cingulate cortex (ACC) is a cortical brain region that has been strongly implicated in processing the affective dimension of noxious input. However, the basic underlying neurochemical mechanisms associated with processing and modulation of noxious information in this brain region remains unknown. Receptor binding assays have revealed a significant concentration of μ-opioid receptors within the region of the ACC. Therefore, in the present experiment morphine, the prototypic μ-opioid receptor agonist, was microinjected into the region of the ACC, and the effect on inflammatory pain using the formalin test was measured. Compared to animals that received vehicle injection and animals that received morphine injection to surrounding cortical areas, morphine into the region of the ACC caused a statistically significant decrease in formalin pain responses. These results confirm previous reports suggesting that the ACC is a critical brain region involved in processing noxious information and point to the importance of the opioid system during activation of the ACC caused by noxious stimulation.

INTRODUCTION

Several lines of evidence indicate that the anterior cingulate cortex (ACC) is an important brain structure for the processing of noxious information. Clinical studies reveal that surgical removal of the ACC and the main projection pathway, the cingulum bundle, results in a significant reduction of malignant and non-malignant chronic pain in humans (Ballantine et al., 1967; Hassenbusch et al., 1990; Santo et al., 1990; Sherma, 1973; Wilkinson et al., 1999; Wong et al., 1997). Psychophysical evidence from patients following the destruction of the ACC and cingulum bundle for the treatment of chronic pain causes a selective decrease in the emotional component of pain, although an alteration in the sensory component of pain perception has also been found (Corkin and Hebben, 1981; Talbot et al., 1995). These results indicate that the ACC receives, integrates, and processes noxious information from both the limbic and somatosensory systems (Coghill, 1994).

Brain imaging studies using the techniques of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have since confirmed that there is an activation of the ACC during acute pain conditions (Coghill, 1994; Derbyshire et al., 1997; Jones et al., 1991). A comparison of regional cerebral blood flow during a chronic pain state and following a successful regional anesthetic block with lidocaine revealed that the ongoing neuropathic pain resulted in activation of the ACC (Hsieh et al., 1995).

In rats, a temporary lesion produced by microinjection of a local anesthetic (lidocaine) into the region of the ACC and cingulum bundle causes a marked decrease in inflammatory pain produced by subcutaneous injection of formalin (Vaccarino and Melzack, 1992). Activation of this region by focal electrical brain stimulation also decreases formalin-induced pain (Fuchs et al., 1996). It was proposed that electrical stimulation activates subcortical regions, such as the periaqueductal grey, to activate the release of descending endogenous endorphins to the spinal cord (Fuchs et al., 1996). Additional work has since shown that electrolytic lesion of the ACC also produces a lower behavioral response in avoidance testing to a noxious stimulus (LaGraize et al., 2000). This finding mimics much of the clinical literature indicating that the ACC modulates the affective component of pain, as avoidance behavior to a noxious stimulus was diminished but the overall degree of pain did not change.

The neurochemical mechanism(s) by which the ACC processes and modulates noxious information remains unclear. Relative to other brain structures, the ACC has an extremely high concentration of opiate receptors. Vogt et al. (1995) suggests that μ-opioid receptors located within the ACC are expressed by efferent pathways to the thalamus and other limbic structures. Therefore, it is possible that destruction and activation of the ACC and main projection pathway decreases pain by altering the normal function of the μ-opioid system. Therefore, the present experiment examined the effect of direct morphine administration to the ACC on an experimental model of inflammatory pain. As morphine is a fairly selective μ-opioid receptor agonist, and the ACC appears to be primarily involved in the emotional aspect of pain processing, we hypothesized that direct microinjections of morphine into the ACC would result in less pain behavior during the prolonged phase of the formalin test.

The formalin test is an inflammatory model of pain that has a characteristic pattern of behavioral responses. There is an initial period of responding that lasts for 5 – 10 min followed by a diminution period. Pain responding re-appears at around 15 – 20 min and gradually increases until 25 – 30 min and then decreases and disappears by 60 – 90 min following the injection. The initial 5-min period of responding is thought to reflect direct activation of nociceptors, whereas the prolonged portion of the formalin test is thought to reflect a significant degree of negative affect. Furthermore, the efficacy of the formalin test has been shown in the testing of narcotic analgesia. (Dubuisson and Dennis, 1977; Corderre et al., 1993; Wheeler-Aceto and Cowan, 1991) Therefore, lower formalin scores should be recorded when morphine is microinjected directly into the ACC as compared to saline microinjection.

METHODS

Subjects

Twenty seven male Sprague-Dawley rats (University of Texas at Arlington vivarium) were used in this study. They were housed in pairs or triplets on a 12:12 light: dark cycle with free access to food and water. The animals were maintained and cared for in accordance with the guidelines established by the International Association for the Study of Pain (Zimmerman, 1983). The Institutional Animal Care and Use Committee at the University of Texas at Arlington approved all experimental protocols.

Anterior Cingulate Cortex Cannula Insertion

The subjects were anaesthetized with a subcutaneous injection of acepromazine (0.65 mg/kg) followed five minutes later with an intramuscular injection of ketamine (50 mg/kg) and xylazine (2.61 mg/kg). Animals were positioned in a stereotaxic frame and with the skull level between bregma and lambda a guide cannula (26-guage) was implanted such that the tip was 0.5 mm above the central region of the ACC (AP, 0.0 mm; L, 0.4 mm; D, 2.6 mm). To prevent clogging of the guide cannula, a stylus (00 insect pin) was inserted and secured in place until the time of testing. Following surgery, animals were isolated until they awoke and were then returned to their normal environment for 7-10 days prior to behavioral testing.

Behavioral Testing

After recovery, the rats were placed into a clean, clear 12" x 12" formalin test cham-

ber and allowed a 30 min habituation period. Animals were removed from the test chamber and administered a subcutaneous injection of 0.5 mg of formalin into the plantar region of the left hindpaw and immediately placed back into the test chamber. After 10 min the stylet was removed and an injection cannula (30-gauge), constructed so that it extended 0.5 mm below the guide cannula, was inserted. Either morphine (10 mg/ml, Sigma) or physiological saline was microinjected using a 1-μl Hamilton syringe attached to the injection cannula via PE 10 tubing. A total of 0.5 μl was injected over a 90 second period and the injection cannula was left in place for an additional 60 seconds to allow for absorption into the surrounding brain region and to minimize spread of the injectate along the injection track. Animals were then returned to the testing chamber to resume behavioral testing for the remaining 25 min following the microinjection procedure.

The experimental pain model employed to induce a long-lasting tonic pain was the formalin test (Dubuisson and Dennis, 1977). Following the injection, each subject was observed and behavior was continuously recorded by determining the amount of time that the subject displayed a prominent behavioral response. The behavioral categories included the amount of time that animals spent with the injected paw down (in contact with the floor of the test chamber), elevated, or licking. For each 5-min time period, a pain score was calculated using a weighted system where: Pain Score = (time with paw elevated + 2 (time licking))/300 sec.

Histology

Following the experimental procedure, animals were sacrificed with an intraperitoneal injection of pentobarbital and perfused with saline followed by 10% formaldehyde. The brains were removed and 80 μm sagittal sections were mounted on glass slides and stained with thionin. Under magnification, the slides were visually examined for cannula tip location. The location of cannula tips were used to designate morphine injected animals into the group of animals that had morphine administered to the ACC (Morphine Hit) or had morphine administered to surrounding cortical tissue (Morphine Miss). All testing was performed blind with respect to drug (morphine or saline) administered and histological analysis was performed blind with respect to behavioral outcome.

RESULTS

Histological analysis revealed that morphine treated animals had cannula tips positioned within or just ventral to the region of the cingulate cortex (Figure 1). Additional animals that received morphine had cannula tip locations that were dorsal and/or lateral to the region of the cingulate cortex. These animals served as controls for the possible diffusion of morphine to adjacent structures. Vehicle injected animals had similar distribution of cannula tip locations as morphine injected animals.

Based on histological analysis of cannula tip location, morphine treated animals were divided into two groups (hit versus miss) and analysis of mean pain scores during the initial 5 min period and the time from 15 – 40 min was performed. Analysis (one-way ANOVA) during the first 5 min time period revealed no significant difference among the groups. The mean pain score for the 15 – 40 minute test period following formalin injection for animals that received vehicle or morphine within or outside of the region of the cingulate cortex is illustrated in Figure 2. As seen, vehicle treated animals displayed the typical response to formalin. Following the initial 0 – 10 min period of responding, a prolonged tonic phase

occurs that is maximum at 25 - 30 min and then begins to gradually decrease. Statistical analysis revealed a significant difference group X time period interaction (p < 0.05). Additional analysis indicated that animals that received morphine within the region of the ACC displayed less pain behaviors during the period from 25 – 40 min post-formalin injection than animals that received either vehicle or morphine to regions outside of the ACC.
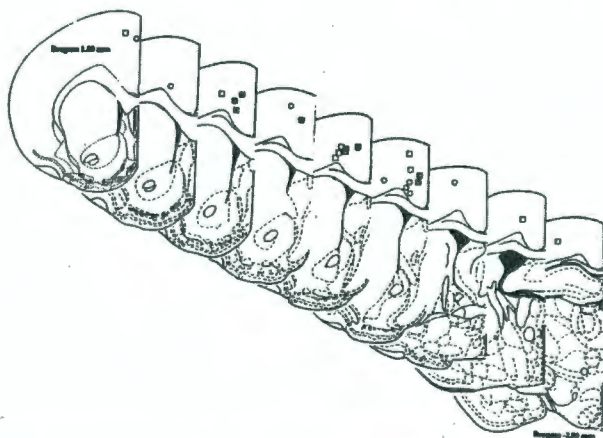

Figure 1: Schematic representation of cannula tip location for animals that received vehicle (○) or morphine outside of the region of the cingulate cortex (■) or within the cingulate cortex (●). Plates are from the atlas of Paxinos and Watson (1986).
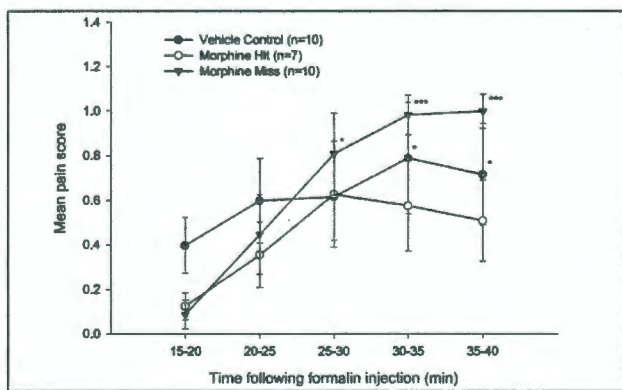
Figure 2: Mean pain scores (±SEM) of animals that received vehicle microinjection (Vehicle Control) or received morphine microinjection within the region of the ACC (Morphine Hit) or surrounding cortical tissue (Morphine Miss). Microinjection of morphine or vehicle occurred between 10 – 15 min. following the formalin injection. Data from the time period 15 – 16 min. following the formalin injection are plotted * p<0.05, *** p<0.001 versus Morphine Hit.

DISCUSSION

The purpose of the present experiment was to investigate the role of the endogenous ACC opioid receptor system in modulating noxious information. The main finding of the present study is that direct administration of morphine within the area of the ACC decreases

the prolonged second phase of the formalin test (15 – 40 min). The difference among the groups during the prolonged portion of the formalin test cannot be accounted for by group differences in sensitivity to formalin since baseline pain scores during the period immediately prior to the microinjections (0 – 5 min) were similar among the groups. Rather, it would appear that morphine activates the endogenous opioid system within the ACC and causes a decrease in pain behaviors. This finding confirms previous clinical, experimental and basic research studies indicating that the ACC is a critical brain structure in processing noxious information.

This is the first report that has examined the neurochemical basis of ACC involvement in pain processing. Although the effect of morphine on the initial period of formalin pain responding was not examined, the time course of effect during the 15 – 40 min test period might reflect a decrease in the affective component of pain processing. Evidence that the primary role of the ACC is in processing the affective component of pain comes from several sources. First, hypnotic suggestion has been shown to affect pain "unpleasantness ratings" when the suggestions were placed prior to receiving identical thermal stimuli (Rainville et al., 1999). Secondly, clinical studies have found a specific decrease in the affective component of the McGill Pain Questionnaire with no alteration in the sensory component (Corkin and Hebben, 1981). The present findings provide a physiological explanation of clinical reports of patients who say that morphine does not decrease the perception of pain, but rather become apathetic to its existence (Price et al., 1980). These positive results lend direct support to our hypothesis that the affective aspect of chronic pain in rats is partially mediated by activity of μ-opioid receptors located within the ACC. Additional studies will examine what role δ-opioid receptors play, but since the δ-opioid receptors do not project to limbic structures, it is unlikely that this system would play a significant role in the affective component of chronic pain.

The function of the ACC in processing of noxious information is extremely complex. It appears paradoxical that both surgical destruction of the ACC and activation of the tissue by focal brain stimulation and morphine produce pain relief. For instance, if the main function of the endogenous opioid system is to decrease pain, then the surgical removal of the ACC should decrease the available source of regulatory control and enhance, rather than decrease, chronic pain. This paradox makes it seem more likely that the ACC is engaging additional neural structures to alter the descending modulatory control system or to alter the input of noxious information at the supraspinal level (Fuchs et al., 1996). Together with the complex neuroanatomical projection system of the ACC (Vogt et al., 1992), it is possible that any of the "downline" structures are affected by altered activity within the ACC. The inherently parallel nature of brain function makes it extremely unlikely that any single structure or system is totally responsible for any one complex sensory experience, and the direct separation of the affective dimension of pain from the discriminative dimension has proven to be very elusive.

Several neuroimaging studies have implicated the ACC in a broad integrative role in the affective and attentional component of pain processing. Using PET scans, Lane et al. (1998) demonstrated that the ACC becomes more active while processing and responding to emotional cues. They suggested that the ACC is partly responsible for the ability to detect emotional cues, including perceived threats to self. Extending this reasoning, it has also been shown that specific individual neurons activate during the anticipation of pain in macaque monkeys (Koyama et al., 1998). In humans, the ACC becomes more active during

the coping behavior associated with the relief of pain (Hsieh et al., 1999). Electrophysiological studies have found that the ACC receives afferent input from both nociceptive specific and wide-dynamic range neurons (Sikes and Vogt, 1992). Assuming that nociceptive specific cells and wide-dynamic range cells encode different information concerning the intensity and location of noxious stimuli, the ACC is in a unique position to process the complex nature of noxious information. Given that affective and motivational behavior is intrinsically bound with limbic structures (Papez, 1937), then altered modulatory action of these structures following ACC lesions and activation would produce a similar behavioral outcome.

From a functional perspective, it is theorized that the ACC motivates pain avoidance behavior by switching awareness. It draws attention to invasive or noxious stimuli in an effort to determine how dramatic the corresponding response should be. This preserves overall energy for the organism while balancing the need for self-protection. This may explain the well-known phenomenon of stress-induced analgesia or episodic analgesia that can occur during traumatic amputation during battle or minor injuries. The extreme threat to self may activate the ACC and engage the release of endogenous opiates. The efferent projections from the ACC to additional limbic system structures would then eliminate the unpleasant feeling associated with the painful condition until either there is no threat to the self or the opiates deplete and break down by normal means.

It must be stressed that the ACC is not the sole system responsible for pain processing but one of many structures and pathways that process this highly complex sensory event. Other examples include the spino-amygdaloid and the spino-reticular pathways that do not project into the ACC. An overall comprehensive view of clinical pain is many years away and will likely include a model for balancing the inputs from the various pathways and systems. This is an inherent feature of many sensory processing systems found in the central nervous system.

ACKNOWLEDGEMENTS

REFERENCES

Ballantine, T., Cassidy, W.L., Flanagas, N.B., and Marino R. 1967. Stereotaxic anterior cingulotomy for neuropsychiatric illness and intractable pain. Journal of Neurosurgery 26: 488-495.

Coderre, T.J., Fundytus, M.E., McKenna, J.E., Dalal, S., and Melzack, R. 1993. The formalin test: a validation of the weighted-scores method of behavioural pain rating. Pain 54: 43-50.

Coghill, R.C., Talbot, J.D., Evans, A.C., Meyer, E., Gjedde, A., Bushnell, M.C., and Duncan, G.H. 1994. Distributed processing of pain and vibration by the human brain. Journal of Neuroscience 14: 4095-4108.

Corkin, S., and Hebben, N. 1981. Subjective estimates of chronic pain before and after psychosurgery or treatment in a pain unit. Pain 1(Suppl.): S150.

Derbyshire, S.W.G., Vogt, B.A., and Jones, A.K.P. 1998. Pain and stroop interference tasks activate separate processing modules in anterior cingulate cortex. Experimental Brain Research 118: 52-60.

Dubuisson, D., and Dennis, S.G. 1977. The formalin test: a quantitative study of the analgesic effects of morphine, meperidine, and brain stimulation of rats and cats. Pain 4: 161-174.

Fuchs, P.N., Balinsky, M., and Melzack, R. 1996. Electrical stimulation of the cingulum bundle and surrounding cortical tissue reduces fromalin-test pain in the rat. Brain Research 743: 116-123.

Hassenbusch, S.J., Pillay, P.K., and Barnett, G.H. 1990. Radiofrequency cingulotomy for intractable cancer pain using stereotaxis guided by magnetic resonance imaging. Neurosurgery 27: 220-223.

Hsieh, J.-C., Belfrage, M., Stone-Elander, S., Hansson, P., and Ingvar M. 1995. Central representation of chronic ongoing neuropathic pain studied by positron emission tomography. Pain 63: 225-236.

Hsieh, J. –C., Stone-Elander, S., and Ingvar, M. 1999. Anticapatory coping of pain expressed in the human anterior cingulate cortex: a positron emission tomography study. Neuroscience Letters 262: 61-64.

Jones, A.K.P., Brown, W.D., Friston, K.J., Qi, L.Y., and Frackowiak, R.S.J. 1991. Cortical and subcortical localization of response to pain in man using positron emission tomography. Proceedings from the London Royal Biological Society 244: 39-44.

Koyama, T., Tanaka, Y.Z., and Mikami, A. 1998. Nociceptive neurons in the macaque anterior cingulated activate during anticipation of pain. Neuroreport 9: 2663-2667.

LaGraize, S.C., LaBuda, C.J., Donahue, R.R., Rutledge, M.A., Jackson, R.L., and Fuchs, P.N. in press. Cingulate cortex lesions attenuate escape/avoidance behavior but have no effect on hyperalgesia following L5 ligation in rats. Society of Neuroscience Abstracts.

Lane, R.D., Reiman, E.M., Axelrod, B., Yun, L.S., Holmes, A., and Schwartz, G.E. 1998. Neural correlates of levels of emotional awareness. Evidence of an interaction between emotion and attention in the anterior cingulate cortex. Journal of Cognitive Neuroscience 10: 525-35.

Mulder, A.H., Hogenboom, F., Wardeh, G., and Schoffelmeer, A.N.M. 1987. Morphine and enkephalins potently inhibit [3H] noradrenaline release from rat brain cortex synaptosomes: Further evidence for a presynaptic localization of m-opioid receptors. Journal of Neurochemistry 48: 1043-1047.

Papez, J.M. 1937. A proposed mechanism of emotion. Archiva Neurologia Psychiatria 38: 723-743.

Paxinos, G., and Watson, C. 1986. The Rat Brain in Stereotaxic Coordinates, 2nd edition, Academic Press, Australia.

Price, D.D., Barrell, J.J., and Gracely, R.H. 1980. A psychophysical analysis of experimental factors that selectively influence the affective dimension of pain. Pain 8: 137-149.

Rainville, P., Carrier, B., Hofbauer, R.K., Bushnell, M.C., and Duncan, G.H. 1999. Dissociation of sensory and affective dimensions of pain using hypnotic modulation. Pain 82: 159-171.

Santo, J.L., Arias, L.M., Barolat, G., Schwartzman, R.J., and Grossman, K. 1990. Bilateral cingulumotomy in the treatment of reflex sympathetic dystrophy. Pain 41: 55-59.

Sherma, T. 1973. Absence of cognitive deficits from bilateral cingulotomy for intractable pain in humans. Texas Medicine 69: 79-82.

Sikes, R.B., and Vogt, B.A. 1992. Nociceptive neurons in area 24 of rabbit cingulate cortex. Journal of Neurophysiology 68: 1720-1732.

Talbot, J.D., Villemure, J.-G., Bushnell, M.C., and Duncan, G.H. 1995. Evaluation of pain perception after anterior capsulotomy: a case report. Somatosensory Motor Research 12: 115-126.

Vaccarino, A.L., and Melzack R. 1992. Temporal processes of formalin pain : differential role of the cingulum bundle, fornix pathway and medial bulboreticular formation. Pain. 49(2): 257-271.

Vogt, B.A., Wiley, G.W., and Jensen, E.L. 1995. Localization of mu and delta opioid receptors to anterior cingulate afferents and projection neurons and input/output model of mu regulation. Experimental Neurology 135: 83-92.

Vogt, L.J., Vogt, B.A., and Sikes, R.W. 1992. Limbic thalamus in rabbit: architecture, projections to cingulate cortex and distribution of muscarinic acetylcholine, GABAA, and opioid receptors. Journal of Comparative Neurology 319: 205-217.

Wheeler-Aceto, H., and Cowan, A. 1991. Standardization of the rat paw formalin test for the evaluation of analgesics. Psychopharmacology 104: 35-44.

Wilkinson, H.A., Davidson, K.M., and Davidson, R.I. 1999. Bilateral anterior cingulotomy for chronic noncancer pain. Neurosurgery 45: 1129-1136.

Wong, E.T., Gunes, S., Gaughan, E., Patt, R.B., Ginsberg, L.E., Hassenbusch, S.J., and Payne, R. 1997. Palliation of intractable cancer pain by MRI-guided cingulotomy. Clinical Journal of Pain 13: 260-263.

Zimmerman, M. 1983. Ethical guidelines for investigators of experimental pain in conscious animals. Pain 16: 109-110.

Effects of Chronic Hypoxia on Cardiovascular and Ventilatory Physiology of Channel Catfish (Ictalurus punctatus)

Philip Silva

Mentor: Dr. Mark Burleson, Department of Biology

ABSTRACT

Ventilatory acclimatization to hypoxia (VAH) is defined as the time-dependent increase in ventilation that occurs with chronic hypoxic exposures (Weil, 1986; Bisgard and Neubauer, 1995). Chronic hypoxia is the reduction of oxygen levels that lasts from several hours to weeks. The effects of chronic hypoxia on cardiovascular and ventilator variables and blood gas and acid/base parameters were examined in conscious channel catfish, *Ictalurus punctatus*. These experiments were designed to determine if channel catfish are able to undergo ventilatory acclimatization response to chronic hypoxia (VAH). The results indicate that chronic hypoxia does cause VAH. This study suggests that VAH in channel catfish is mediated by an increase in the sensitivity of the oxygen-sensitive chemoreceptors.

INTRODUCTION

The ability to sense and respond to fluctuating levels of oxygen in the environment is one of the most essential physiological mechanisms for survival. Most vertebrates can only survive for a few minutes without oxygen; thus, the ability to adapt to minimal oxygen levels in the environment is critical. Branchial oxygen-sensitive chemoreceptors exert dominant control over cardioventilatory reflexes in fish. Hypoxia is the decrease of oxygen levels both internal and external. Hypoxia triggers reflex responses in these oxygen-sensitive chemoreceptors. How certain vertebrates deal with a decrease in oxygen (and adapt to it) is a topic of great environmental and clinical importance. The effects of chronic hypoxia on cardioventilatory control have been studied in mammals and birds, but have yet to be studied in lower vertebrates. Nothing is known about how chronic hypoxia affects cardioventilatory responses in fish and this represents a critical gap in physiological knowledge and understanding.

Fish make the most appropriate models for the study of the effects of chronic long-term hypoxia because the aquatic environment is very unstable and fish must be able to adapt to the constantly changing oxygen content in aquatic habitats. With the reduction of photosynthesis in the evening, ponds and other small isolated bodies of water tend to lose oxygen, causing frequent hypoxic conditions. Unlike air-breathing animals, who are almost never exposed to long terms of hypoxia unless in a high altitude environment, fish are more frequently exposed to hypoxia. This experiment is designed to observe how the channel catfish, *Ictalarus punctatus*, reacts to chronic hypoxia and if it undergoes ventilatory acclimatization to hypoxia. Cardioventilatory variables will be observed and recorded to verify the effects of chronic hypoxia through the use of various experimental methods.

HYPOTHESIS

Chronic hypoxia will alter the cardioventilatory response of the channel catfish and it will undergo acclimatization.

ANIMAL PREPARATION

Channel catfish were obtained from a commercial supplier and maintained in filtered, dechlorinated tap water in a temperature-controlled room (19±°C), on 12/12 light/

dark cycles for at least one week before experimentation (Burleson and Smatresk, 1999).

Surgery

Fish were anesthetized in MS 222 (tricaine methanesulphonate, 0.01% solution) dissolved in water bubbled with 100% oxygen and artificially ventilated on a surgery table. An opercular cannula was inserted by drilling a hole in the operculum and guiding a PE160 catheter on the underside of the operculum, tightening with a cuff and sleeve attachment tied with a string to secure the cannula. The dorsal aorta was cannulated by making a small hole in the roof of the mouth posterior to the last gill arch and inserting PE 50, which was secured with silk suture and led out a small hole in the snout and secured with a cuff and sleeve assembly. Arterial cannula was filled with heparinized Cortland saline and plugged (Wolf, 1963).

PROTOCOL

After surgery the catfish were maintained in twenty-gallon aquaria for six days. Two aquaria were maintained at 50% (approximately 75 torr) atmospheric oxygen concentration (experimental) and two were maintained at 100% (approximately 150 torr) atmospheric oxygen. After six days individual catfish were moved to an experimental chamber in the laboratory. The fish were allowed to adjust in a darkened room with minimal noise at the acclimation oxygen level. After a 24-hour adjustment period, the dorsal aorta and opercular cannulae were attached to pressure transducers (Validyne CD15 Carrier Demodulator) and the recording apparatus (Coulbourn Instruments Transducer Coupler Type A Strain Gage Bridge), and the fish (n=8) were allowed to recover until they displayed stable ventilation for at least 30 minutes. The fish were exposed to normoxia (air saturation) and two levels of hypoxia (approximately 50% and 30%). Oxygen content in the water was controlled using a gas mixing flow meter (Cameron Instruments). For each oxygen level a three-minute recording was done to measure cardiovascular and ventilatory variables, and a blood sample was drawn from the DA cannula at the end of the recording period for measurement of blood gas, acid-base, and hematocrit. The fish were killed at the end of the experiment by an excess of anesthetic (0.05% MS 222) injected into the DA catheter. Comparative analysis will be done using an ANOVA graph comparing the results between the control and the experimental fish and a Scheffé test for post-hoc comparisons.

Blood gas and acid-base measurement

Blood samples (~ 0.5 ml) were taken from the arterial cannula immediately after the recording period. Oxygen partial pressure (Po_2) and pH were analyzed using a blood gas meter (Cameron Instruments). Total carbon dioxide (Cco_2, mmol/l) was analyzed using a Capni-Con (Cameron Instruments). Arterial oxygen (mmol/l) content was measured using the Oxycon blood oxygen content analyzer (Cameron Instruments). Hematocrit was calculated as percent packed red cell volume after three minutes of centrifugation (Burleson and Smatresk 1999).

Cardiovascular and ventilatory measurements

Heart rate (fH, beats/min), dorsal aorta blood pressure (P_{DA}, mmHg), gill ventilation frequency (fG, beats/min), and opercular pressure amplitude (P_{OP}, mmHg) were measured in the experiment. Heart rate and gill ventilation frequency were measured in 15-second intervals. Data were statistically analyzed with a two-way repeated measured ANOVA and least signifi-

cant difference post-hoc comparisons using various commercial software programs (Statistica). Graphs were prepared using a different commercial software program (Origin).

RESULTS

Gill ventilation rate (fG) increased in both the normoxic and hypoxic acclimatized fish. Gill ventilation rate was significantly higher in the normoxic fish through all three oxygen level tensions in the water. The largest change occurred between the transition from normoxia (148 torr) to hypoxia 1 (75 torr) in both groups of fish.
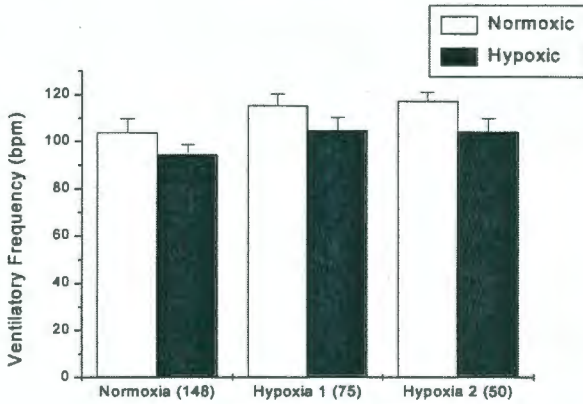

Figure 1: Gill Ventilation Rate (fG).
Gill ventilation rate (fG, beats/min) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

Opercular presser amplitude (P_{OP}) levels rose in both groups of fish in each oxygen level. As oxygen tension was decreased opercular pressure increased in both groups. The hypoxic fish showed a greater increase in opercular pressure over the normoxic control group.
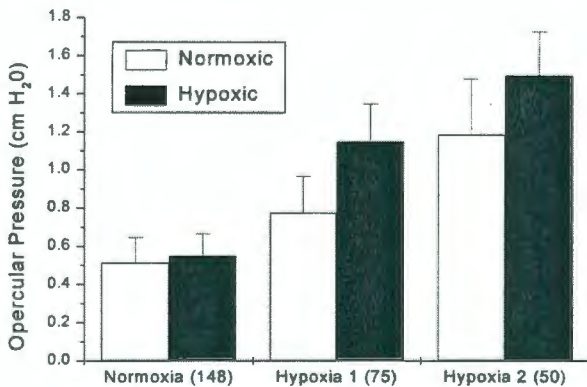
Figure 2: Opercular Pressure Amplitude (P_{OP}).
Opercular pressure amplitude (P_{OP}, cm H_2O) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish

Heart rate decreased as oxygen levels decreased in both groups. Although both groups showed decrease, the chronic hypoxic fish had an overall lower heart rate. The heart rate in the normoxic fish was at very similar levels at all three levels of oxygen tension. There was not a significant amount of fluctuation in the heart rate of the normoxic fish.
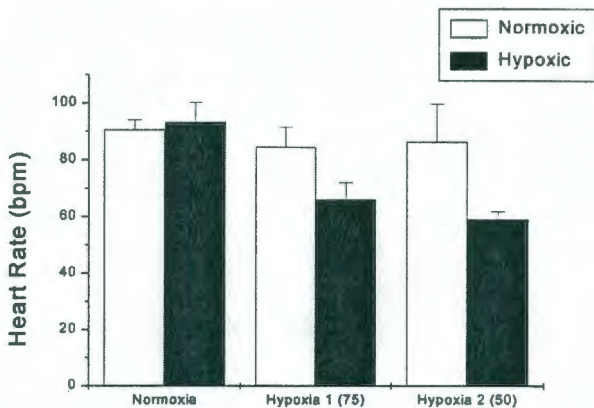
Figure 3: Heart Rate (fH).
Heart rate (fH. beats/min), during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

Partial Pressure decreased significantly in both chronically hypoxic and normoxic fish. There was no significant decrease between the normoxic and hypoxic fish in the three levels of oxygen tension. As the oxygen tension levels of the water were decreased so was the partial pressure of oxygen in both sets of fish.
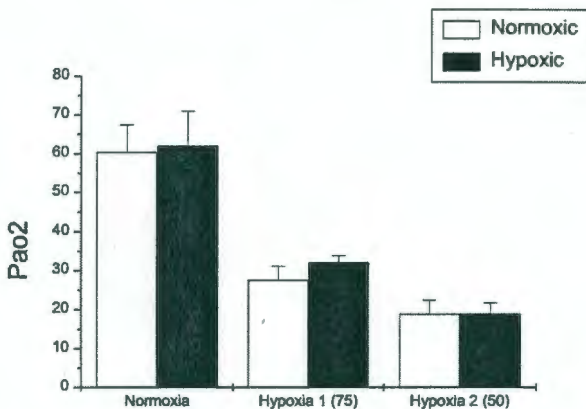
Figure 4: Arterial Partial Pressure of Oxygen (PaO_2).
Arterial partial pressure of oxygen (PaO_2) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

There was a significant decrease in all three levels of oxygen. The content decreased as the oxygen tension decreased. Arterial carbon dioxide content decreased significantly in both sets of fish, although the chronically hypoxic group had a lower carbon dioxide content than the normoxic grouping all three oxygen tensions.
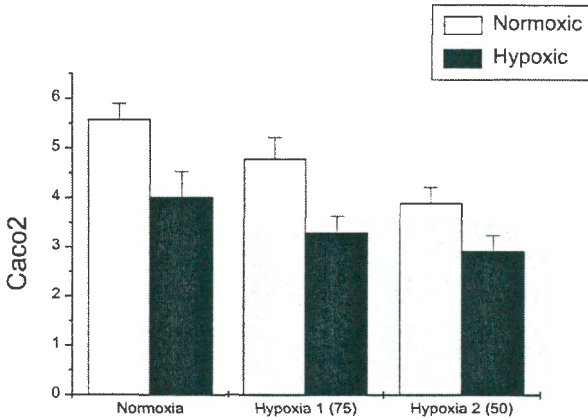
Figure 5: Arterial Carbon Dioxide Content ($CaCO_2$).
Arterial carbon dioxide content ($CaCO_2$ mmol/l) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

The pH level was similar in both sets of fish and did not show any fluctuation at any one of the three levels of oxygen.
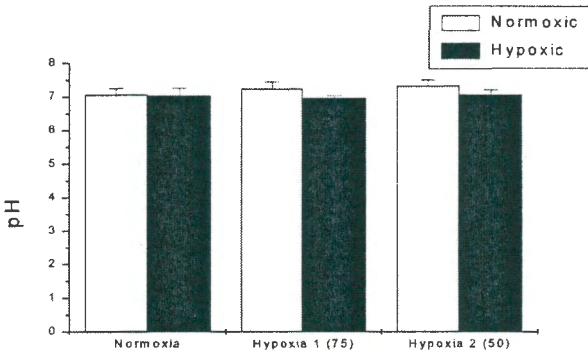
Figure 6: pH Levels.
pH during normoxia, hypoxia 1 (75 torr),
and hypoxia 2 (50 torr) in conscious catfish.

In the level of normoxia the normoxic fish showed greater arterial oxygen content. Both sets of fish demonstrated similar levels of oxygen content in hypoxia 1, but in hypoxia 2 the chronically hypoxic fish showed a greater content of arterial oxygen.
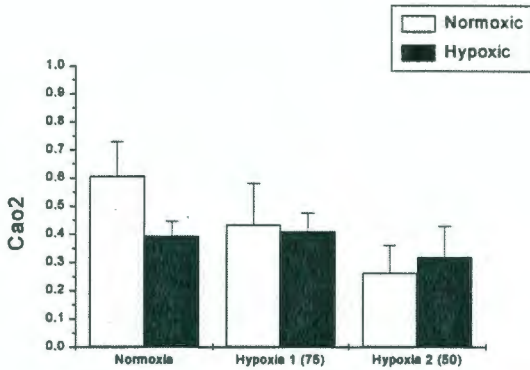
Figure 7: Arterial oxygen content (CaO$_2$).
Arterial oxygen content (CaO$_2$, mmol/l) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

Both the chronically hypoxic and normoxic control fish were kept in similar oxygen
content when the experiments were being performed and data collected.
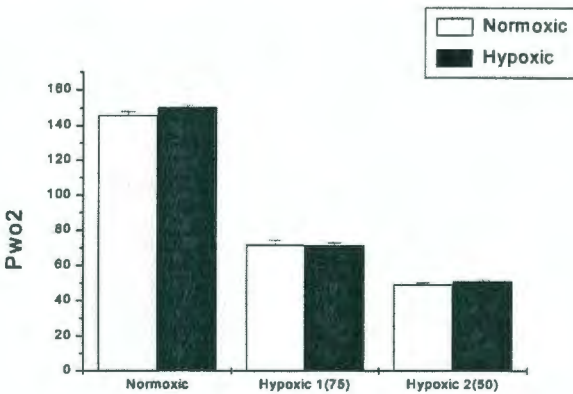
Figure 8: Oxygen Content in the Water (PwO$_2$).
Oxygen content in water (PwO$_2$, torr) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

In the level of normoxia both sets of fish showed very similar levels of hematocrit,
but in both levels of hypoxia the chronically hypoxic fish showed a higher hematocrit count.
As hypoxia was increased the hematocrit count was increased in both sets of fish.
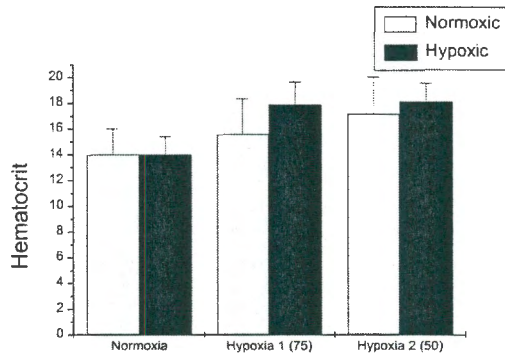
Figure 9: Hematocrit Count.
Hematocrit percent packed red cell volume (%) during normoxia,
hypoxia 1 (75 torr), and hypoxia 2 (50 torr) in conscious catfish.

DISCUSSION

Hypoxia is a reduction in the oxygen level. Hypoxia is caused by various biotic and abiotic factors in the environment. Small ponds, tropical marshes, or freshwater lakes experience hypoxia due to lack of photosynthesis in the evening, causing oxygen levels to greatly reduce. In the air environment, high altitude environments, such as mountains, have low oxygen levels and remain hypoxic.

Hypoxia causes a number of cardiovascular and ventilatory reflex responses in vertebrates. A reduction in oxygen in the environment impacts the amount of oxygen present in the blood, causing a decrease in oxygen pressure. This causes stimulation to the chemoreceptors, which triggers an increase in ventilation. Hypoxia also causes most vertebrates to increase the red blood cell count and also the blood hemoglobin content. Raising the red blood cell and hemoglobin levels allows a greater capacity of oxygen in the blood. There are three levels of hypoxia: acute hypoxia (AH), exposure lasting from minutes up to an hour; "sustained hypoxia" (SH), lasting hours; and chronic hypoxia (CH), which lasts weeks or longer (M.A. Hanson 1998).

Studies indicate that there are at least two groups of oxygen-sensitive chemoreceptors in the gills of fishes. Aquatic (external) hypoxia stimulates both an increase in ventilation and a decrease in heart rate. Internal hypoxia (hypoxemia) stimulates ventilation, but has no significant effect on heart rate. Therefore, there seems to be one group of chemoreceptors monitoring external oxygen levels and another group monitoring internal oxygen levels. The response time of fish to external hypoxia is very short, 5-7 seconds. This strongly suggests that the oxygen receptor initiates the change of cardiovascular and ventilatory adjustments before the oxygen reaches metabolically active tissue. This supports the idea that there are separate oxygen chemoreceptors for internal and external hypoxia. Internal receptors respond to internal hypoxia even if hypoxia is absent in the aquatic environment.

Hypoxemia in the absence of change in oxygen in the aquatic atmosphere appears to be common in wild fish. This is due to blood loss from injury or parasitism. Chemical pollutants, such as nitrite (NO_2^-) and nitrate (NO_3^-) also cause hypoxemia in wild fish. The

nitrite combines with the hemoglobin of the fish forming methemogloin, turning the blood a brown color and blocking the ability of the hemoglobin to bind with oxygen. This condition is commonly called "brown blood disease" because of the excess of nitrite turning the blood a brown color. Having two groups of oxygen sensitive chemoreceptors enables fish to respond differently to conditions of hypoxia and hypoxemia in order to maintain oxygen uptake (Burleson 1995).

Acute Hypoxia

Nearly all of the research has focused mainly on acute hypoxia. During exposure to acute hypoxia, blood pressure significantly increased as oxygen tension in the water fell. Heart rate decreased during exposure to acute hypoxia as the oxygen tension decreased. Opercular pressure and gill ventilation rose significantly during acute hypoxia. This was due to the reflex response in the oxygen-sensitive chemoreceptors (Burleson and Smatresk 1989).

Chronic Hypoxia

There have been only two other experiments using (carp and flounder) prior to this one in which chronic hypoxia has been studied (Lomholt and Johansen 1978,1978). Blood/gas measurements were not taken in these experiments. Lomholt and Johansen focused on ventilation. Carp held at 30 torr for four weeks did not increase ventilation and, therefore, did not undergo ventilatory acclimatization to hypoxia (VAH). It is suggested that the carp experienced hypoxic desensitization (HD), which takes effect over the course of the month during which the oxygen-sensitive chemoreceptors become desensitized. Lomholt and Johansen in a later experiment studied chronic hypoxia in flounder. One flounder was kept chronically hypoxic at 30 torr for three weeks, and this resulted in the flounder undergoing ventilation increase, suggesting VAH.

Ventilation

The most recent experiment on chronic hypoxia, as described in this paper, confirms that channel catfish undergo ventilatory acclimatization to hypoxia (VAH). The data shows that the chronic hypoxic catfish had greater opercular pressure amplitude than the normoxic control, showing greater sensitivity in the oxygen chemoreceptors which indicates acclimatization of the chronically hypoxic catfish.

Heart Rate

Heart rate was also lower in the chronically hypoxic fish due to acclimatization, indicating a reflex response triggered by hypoxia. The heart rate lowers so that the resonance time that the blood stays in the gills is longer, increasing oxygen uptake in the blood in order to acquire more oxygen from the water.

Hematocrit

An increased hematocrit count in both sets of fish was observed, but the chronically hypoxic fish had an overall greater hematocrit count due to a hypoxic reflex response in order to increase oxygen uptake. This was also evident in Rainbow Trout (Tetens and Lykkeboe1981).

Arterial Carbon Dioxide Content

In both sets of fish, carbon dioxide content decreased significantly with decreasing oxygen tension in the water. This indicates increased oxygen uptake causing more carbon dioxide blow off. The carbon dioxide content in the normoxic fish was significantly higher than in the hypoxia acclimated fish, indicating that the hypoxic acclimated fish had greater ventilation due to more sensitive oxygen chemoreceptors.

Arterial Oxygen Content

There is no significant difference in arterial oxygen content between both sets of fish. This may have been due to experimental troubles, such as holes in the data due to failed cannulae and to the machine being repaired.

Arterial Partial Pressure of Oxygen

There is a significant decrease in both the normoxic and hypoxic groups in all three levels of oxygen tension. The arterial partial pressure in both sets of fish seemed to have decreased to the same amount in all three levels of oxygen tension. This is an important observation because it shows that the hypoxic acclimated fish have more sensitive oxygen-sensitive chemoreceptors due to opercular pressure in the hypoxic fish being significantly higher than in the normoxic group, although there is no significance in partial pressure between the two groups.

Oxygen Tension of the Water

There is no significant difference in the oxygen tensions between both sets of fish. This is important in the data collecting process. Both sets underwent almost identical levels of oxygen tension in order to collect the most accurate data possible in the experiment.

Potential for Hydrogen (pH) Levels

The pH levels in both the normoxic and hypoxic fish were very similar in all three levels of oxygen tension. The pH levels were also very similar in both groups; this indicates that the fish were maintained at minimal stress levels. If pH was decreased, this would signify an increase in lactic acid buildup due to stress. This was not seen in the pH level in either group of fish, indicating that the fish were at a low stress level.

Like most mammals, catfish also demonstrate a physiological acclimatization due to chronic hypoxia. These processes appear to be mediated by oxygen sensitivity in the oxygen-sensitive chemoreceptors. This is evidence for evolution of response because it is widespread through vertebrates and also illustrates the utility of channel catfish as the prime experimental model.

ACKNOWLEDGEMENTS

REFERENCES

1.   Burleson, M.L., and N.J. Smatresk. The effect of decerebration and anesthesia on the reflex responses to hypoxia in catfish. Canadian Journal of Zoology. 1989; 67: 630-635.

2. Burleson, M.L. Oxygen availability: Sensory Systems. Biochemistry and Molecular Biology of Fishes, vol. 5, edited by P.W. Hochachka and T.P. Mommsen. Elsevier. 1995; 1-17.

3. Hanson, M.A. Role of chemoreceptors in effects of chronic hypoxia. Comparative Biochemistry and Physiology, Vol. 119A, 1997; 3: 695-703.

4. Lomholt, J.P., and Johansen, K. Hypoxia acclimation in carp-How it affects O_2 uptake, ventilation and O_2 extraction from water. Physiology & Zoology. 1979; 52: 38-49.

5. Lomholt, J.P., and Johansen, K. The ventilation, extraction and uptake of oxygen in undisturbed flounders, *Platichthys flesus*: Responses to hypoxia acclimation. Journal of Experimental Biology 1979; 83: 169-170.

6. Powell, F.L. and Milsom, W.K. Time domains of the hypoxic ventilatory response. Respiratory Physiology 1998; 112: 123-134.

7. Tetens, Vilhelm and Lykkeboe, Gunnar. Blood Respiratory Properties of Rainbow Trout, *Salmo gairdneri*: Responses to hypoxia acclimation and anoxic incubation of blood in vitro. Journal of Comparative Physiology 1981; 145: 117-125.

Factors Associated with Student Performance in
The University of Texas at Arlington
Mathematics Courses in Fall 1999 and Spring 2000

Ni Khoi Tri

Mentor: Dr. D.L. Hawkins, Department of Mathematics

ABSTRACT

This study examines the relationships between pretest scores, attendance, gender, ethnicity and student performance in The University of Texas at Arlington Analytic Geometry and Calculus I classes. Performance is measured in terms of final exam scores, A/B rates, and drop/fail rates. It is found that pretest scores are a stronger predictor of performance than attendance and that no gender differences exist in Analytic Geometry and Calculus I. Also, it is seen that international students score higher in these courses than students of other ethnicities.

INTRODUCTION

In recent years, The University of Texas at Arlington Mathematics Department has experienced high drop/fail rates in Analytic Geometry (MATH 1325) and Calculus I (MATH 1426). To identify factors associated with this problem, as well as factors associated with student success, data was collected in Fall 1999 and Spring 2000. Three performance measures were considered: average final exam scores, A/B rates, and drop/fail rates. This study focuses only on students taking these classes for the first time. We examine:

(1) the relationship between pretest scores and the performance measures

(2) the relationship between attendance and the performance measures

(3) which of pretest scores or attendance is the stronger predictor of the performance measures

(4) the issue of ethnic and gender stereotypes. Is there evidence that Asians outperform other ethnic groups or that men's performance surpasses that of women in these classes?

DATA SOURCES

Pretest Data: In both courses, students took a pretest on the first class day, which measured knowledge of course pre-requisites. The Analytic Geometry pretest consisted of algebra and trigonometry questions, while the pretest for Calculus I consisted of pre-calculus questions developed by the Mathematical Association of America. Students who did not attend the first day of class were asked to go by the Mathematics Department to complete the pretest. The response rate for this test was about 90% for each course.

Attendance Data: This data was collected using MAV EXPRESS card readers[1]. Ideally, attendance registration was required for entrance into the classroom. In actuality, attendance was not always recorded; that is, a small percentage of students refused to participate. Such students have attendance percentages of, or very close to, zero in this data, and tend to fall into the lowest (< 20%) attendance category in the ensuing graphs.

---

1 The MAV EXPRESS card is The University of Texas at Arlington student's permanent identification card. The MAV EXPRESS card reader is a device used to read the card's magnetic strip, encoded with the student's identification number.

Registrar Data: At the end of each semester, the registrar provided information which included: student course grades, course withdrawls, gender, age, and ethnicity.

Compiling the Data: All three data sources were merged using student ID numbers and placed into a SAS database for statistical analysis.

VARIABLE DEFINITIONS

The following section defines variables used in this report:

Final Exam: Percentage of the final exam answered correctly. A standardized final exam was given to all sections of each course.

A/B Rate: Percentage of students who received a final course grade of A or B.

Drop/Fail Rate: Percentage of students who dropped or failed their math class.

Pretest: Percentage of the pretest answered correctly.

Attendance: Percentage of the lectures a student attended, up until his or her drop date, if he or she dropped.

Gender: Student's gender.

Ethnicity: Each student is categorized into one of the following six categories: Black, Indian, White, Hispanic, Asian, and Non-Residential Alien. Indians were excluded in this study due to their extremely sparse numbers.

REPORT STRUCTURE

The structure of the report is as follows:

   (1)   Results for Analytic Geometry (Figures and Discussion)
   (2)   Results for Calculus I (Figures and Discussion)
   (3)   Conclusions

For each of the first two sections, a series of bar graphs and tables are presented, followed by discussion. Atop each bar in the bar graphs, is an "error bar" that represents one standard error. The longer this bar, the less precise (statistically) is the corresponding estimate. The tables show the distributions of the particular factor variable being studied in the corresponding bar graph. The conclusions section follows the format of the questions in the introduction.

Pretest(%)	N	%
< 20	11	5
20 - 39	72	34
40 - 59	84	39
60 - 79	42	20
> 79	4	2
Fig. 1.1.1	Corr:	0.35
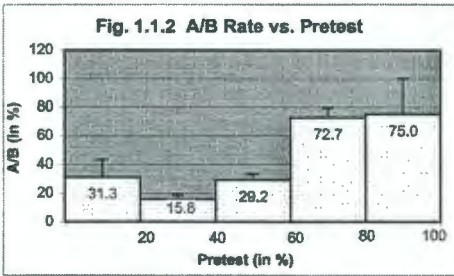
Pretest(%)	N	%
< 20	16	5
20 - 39	116	39
40 - 59	118	40
60 - 79	44	15
> 79	4	1
Fig. 1.1.2	Corr:	0.34
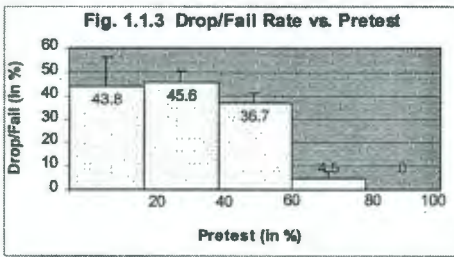
Pretest(%)	N	%
< 20	16	5
20 - 39	116	39
40 - 59	118	40
60 - 79	44	15
> 79	4	1
Fig. 1.1.3	Corr:	-0.26

Figures 1.1.1 – 1.1.3
**Pretest Scores versus Average Final Exam Score, A/B Rate, and
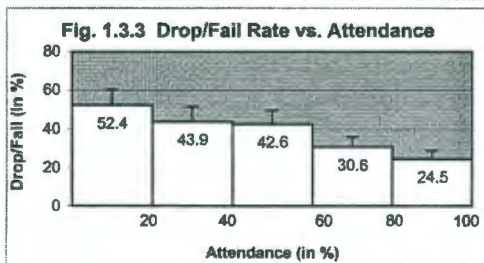Drop/Fail Rate in Analytic Geometry (Fall 1999)**

Fig. 1.2.1 Final Exam vs. Pretest
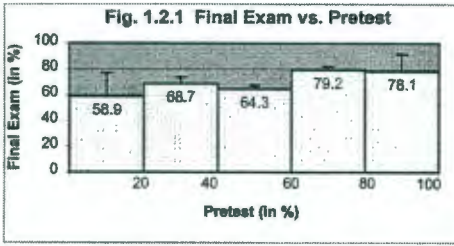
Pretest(%)	N	%
< 20	3	4
20 - 39	10	12
40 - 59	37	44
60 - 79	28	33
> 79	7	8

Fig. 1.2.1   Corr: 0.28

Fig. 1.2.2 A/B Rate vs. Pretest
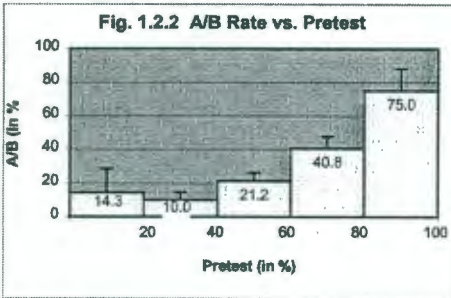
Pretest(%)	N	%
< 20	7	4
20 - 39	40	23
40 - 59	65	38
60 - 79	49	28
> 79	12	7

Fig. 1.2.2   Corr: 0.36

Fig. 1.2.3 Drop/Fail Rate vs. Pretest

Pretest(%)	N	%
< 20	7	4
20 - 39	40	23
40 - 59	65	38
80 - 79	49	28
> 79	12	7

Fig. 1.2.3   Corr: -0.30

Figures 1.2.1 – 1.2.3
Pretest Scores versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Analytic Geometry (Spring 2000)

Fig. 1.3.1 Final Exam vs. Attendance

Final Exam (in %) / Attendance (in %)

Bar values: 64.9, 66.8, 60.4, 69.5, 69.3
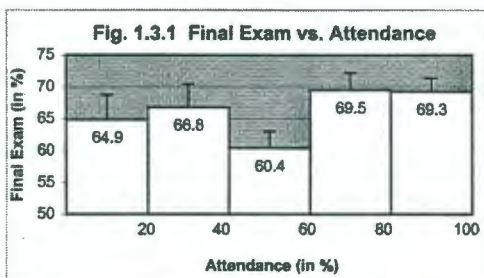
Attn(%)	N	%
< 20	24	11
20 - 39	25	11
40 - 59	34	16
60 - 79	55	25
> 79	80	37
Fig. 1.3.1	Corr:	0.11

Fig. 1.3.2 A/B Rate vs. Attendance

A/B (in %) / Attendance (in %)

Bar values: 16.7, 29.3, 17.0, 38.9, 40.2
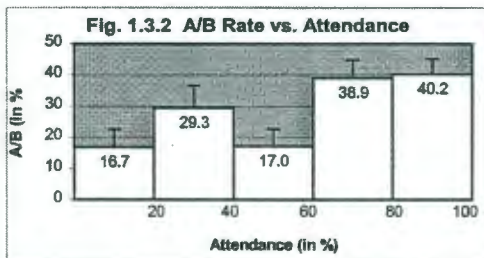
Attn(%)	N	%
< 20	42	15
20 - 39	41	15
40 - 59	48	17
60 - 79	72	26
> 79	77	28
Fig. 1.3.2	Corr:	0.18

Fig. 1.3.3 Drop/Fail Rate vs. Attendance

Drop/Fail (in %) / Attendance (in %)

Bar values: 52.4, 43.9, 42.6, 30.6, 24.5

Attn(%)	N	%
< 20	42	15
20 - 39	41	15
40 - 59	48	17
60 - 79	72	26
> 79	77	28
Fig. 1.3.3	Corr:	-0.21

Figures 1.3.1 – 1.3.3
Attendance versus Average Final Exam Score, A/B Rate, and
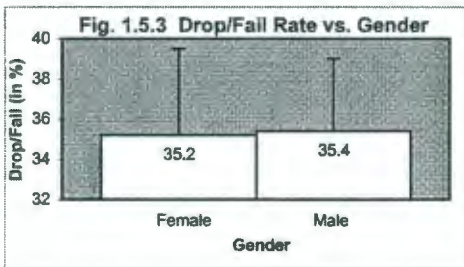**Drop/Fail Rate in Analytic Geometry (Fall 1999)**

Fig. 1.4.1 Final Exam vs. Attendance
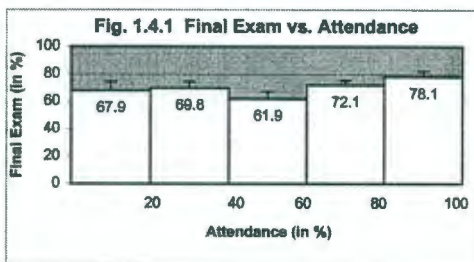
Attn(%)	N	%
< 20	34	20
20 - 39	33	19
40 - 59	35	20
60 - 79	33	19
> 79	39	22

Fig. 1.4.1 Corr: 0.20

Fig. 1.4.2 A/B Rate vs. Attendance
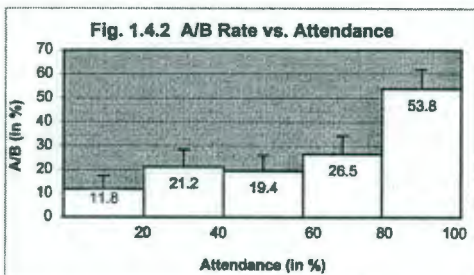
Attn(%)	N	%
< 20	11	13
20 - 39	14	16
40 - 59	19	22
60 - 79	16	19
> 79	25	29

Fig. 1.4.2 Corr: 0.30

Fig. 1.4.3 Drop/Fail Rate vs. Attendance
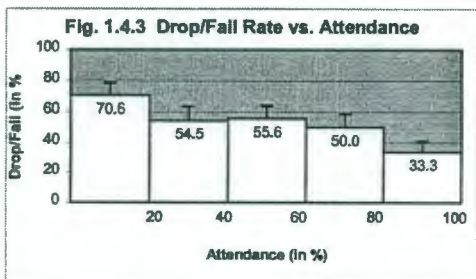
Attn(%)	N	%
< 20	11	13
20 - 39	14	16
40 - 59	19	22
60 - 79	16	19
> 79	25	29

Fig. 1.4.3 Corr: -0.23

Figures 1.4.1 – 1.4.3
Attendance versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Analytic Geometry (Spring 2000)
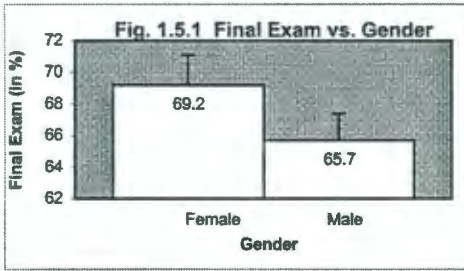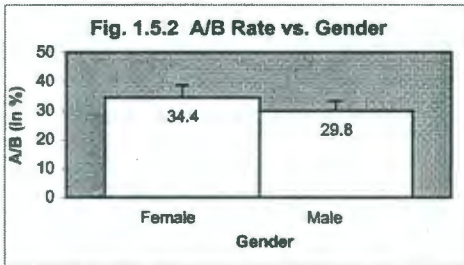
Figures 1.5.1 – 1.5.3
Gender versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Analytic Geometry (Fall 1999)
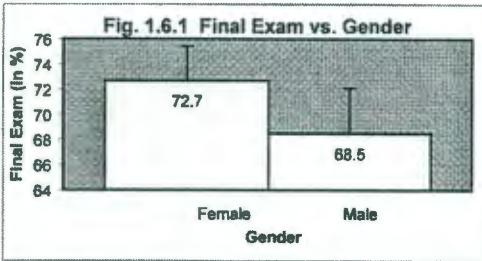
Gender	N	%
Female	51	61
Male	33	39

Fig. 1.6.1
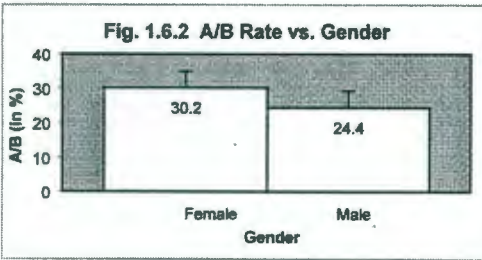
Gender	N	%
Female	97	55
Male	78	45

Fig. 1.6.2
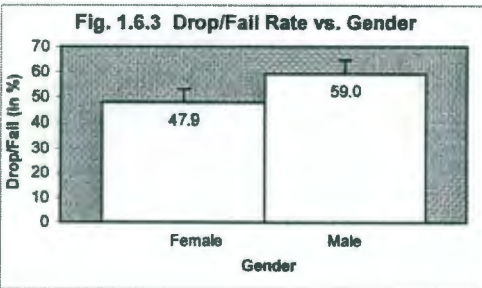
Gender	N	%
Female	97	55
Male	78	45

Fig. 1.6.3

Figures 1.6.1 – 1.6.3
Gender versus Average Final Exam Score, A/B Rate, and
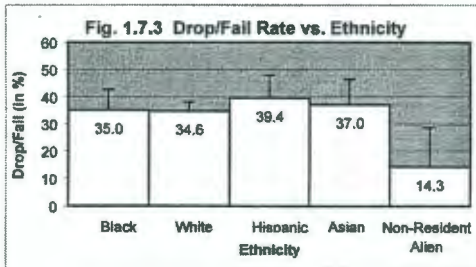**Drop/Fail Rate in Analytic Geometry (Spring 2000)**

Fig. 1.7.1 Final Exam vs. Ethnicity

Black	White	Hispanic	Asian	Non-Resident Alien
61.0	66.9	66.8	74.6	89.4
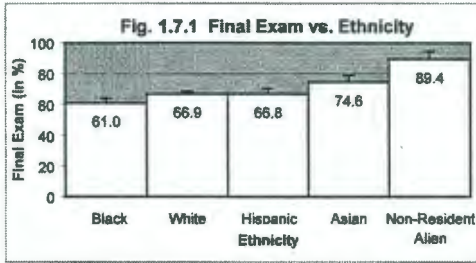
Ethnic	N	%
Black	31	15
White	131	62
Hispanic	24	11
Asian	19	9
Non-Res.	6	3
Fig. 1.7.1		
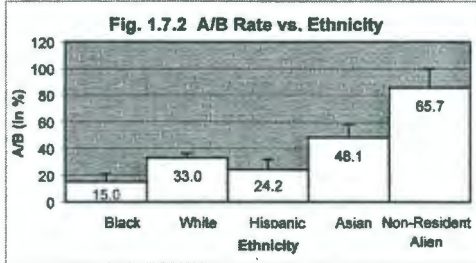
Fig. 1.7.2 A/B Rate vs. Ethnicity

Black	White	Hispanic	Asian	Non-Resident Alien
15.0	33.0	24.2	48.1	65.7

Ethnic	N	%
Black	40	13
White	192	64
Hispanic	33	11
Asian	27	9
Non-Res.	7	2
Fig. 1.7.2		

Fig. 1.7.3 Drop/Fail Rate vs. Ethnicity

Black	White	Hispanic	Asian	Non-Resident Alien
35.0	34.6	39.4	37.0	14.3

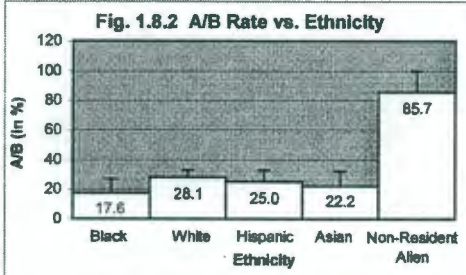| Ethnic	N	%
Black	40	13
White	192	64
Hispanic	33	11
Asian	27	9
Non-Res.	7	2
Fig. 1.7.3		

Figures 1.7.1 – 1.7.3
Ethnicity versus Average Final Exam Score, A/B Rate, and
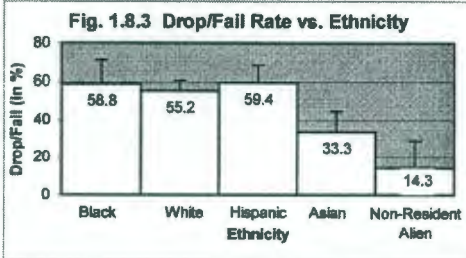**Drop/Fail Rate in Analytic Geometry (Fall 1999)**

Fig. 1.8.1 Final Exam vs. Ethnicity
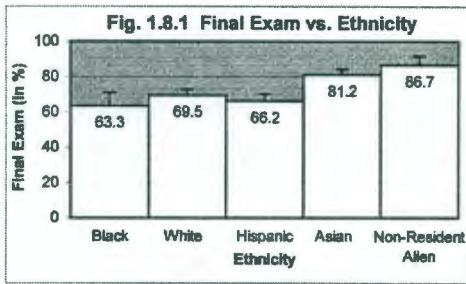
Ethnic	N	%
Black	8	10
White	45	54
Hispanic	14	17
Asian	11	13
Non-Res.	5	6

Fig. 1.8.1

Fig. 1.8.2 A/B Rate vs. Ethnicity

Ethnic	N	%
Black	17	10
White	97	57
Hispanic	32	19
Asian	18	11
Non-Res.	7	4

Fig. 1.8.2

Fig. 1.8.3 Drop/Fail Rate vs. Ethnicity

Ethnic	N	%
Black	17	10
White	97	57
Hispanic	32	19
Asian	18	11
Non-Res.	7	4

Fig. 1.8.3

Figures 1.8.1 – 1.8.3
Ethnicity versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Analytic Geometry (Spring 2000)

ANALYTIC GEOMETRY DISCUSSION

Before examining the relationships with the factor variables, we first examine the distributions of these variables. In the table next to Fig. 1.1.2, we see that only 16% of the students scored 60% or higher on the pretest coming into Analytic Geometry in Fall. Although the table for Fig. 1.2.2 shows this percentage to be 35% in the Spring, less than half the students passed the pretest covering material important for succeeding in this course.

Also important for success is attendance. The tables next to Fig. 1.3.2 and 1.4.2 show that only 54% of the students in Fall and 48% of the students in Spring attended 60% or more of the lectures. It is important to note that the numbers for the < 20% attendance category are misleading as they contain the students who did not register their attendance.

Comparing the pretest correlations in Fig. 1.1 and 1.2 with the corresponding attendance correlations in Fig. 1.3 and 1.4, we see that pretest correlations are stronger than the

attendance correlations. Thus stronger pretest is a stronger predictor of performance than is attendance.

Final Exam

All the tables for final exam have smaller N's than the corresponding tables for A/B and drop/fail. This decrease in the number of students reflects those who dropped before the final exam. Fig. 1.1.1 and 1.2.1 show that final exam scores averaged about 15% higher for students scoring 60% or better on the pretest than for those who did not. However, within the below 60% pretest categories, and within the above 60% pretest categories, there is no clear association between final exam and pretest.

Unlike pretest scores, attendance is only very weakly associated with final exam. Fig. 1.3.1 shows that in Fall, final exam scores are in the 60% range for all levels of attendance. For Spring, Fig. 1.4.1 also shows only slightly increasing final exam scores as attendance category improves.

There seems to be a weak relationship between final exam and gender. In Fig. 1.5.1, females on average scored 3.5 points higher than males, while in Fig. 1.6.1 they scored 4.2 points higher. However, the difference is not statistically significant in either case. Males comprised 59% of the students in the Fall (see table for Fig. 1.5.1) and 39% in the Spring (see table for Fig. 1.6.1).

It seems that students of different ethnic groups scored differently on the final. The non-residential aliens outperformed all other groups. In Fig. 1.7.1 (Fall), they averaged 89.4% on the final and 86.7% in Fig. 1.8.1 (Spring). In these two figures we also see that Asians performed second best and that the Blacks averaged the worst.

A/B Rate

In addition to the average final exam score, A/B rate is another measure of student success. When we compare pretest scores to A/B rate in Fig. 1.1.2 and 1.2.2, we see, as we would expect, that as pretest scores increase, so does the A/B rate.

This relationship between A/B rate and pretest is present between A/B rate and attendance as well. Although Fig. 1.3.2 does not show a smooth trend, still overall higher attendance means a student is more likely to attain an A or B. Fig. 1.4.2 also shows that in general better attendance means higher A/B rate.

Following the graphs for attendance are graphs concerning gender. It appears in Fig. 1.5.2 and 1.6.2 that females have slightly higher A/B rates than males in Analytic Geometry. On average their A/B rate is 4.6% higher than males in Fall and 5.8% higher in Spring. However, neither difference is statistically significant.

For A/B rates we again see differences across ethnic groups. Non-residential aliens have the highest A/B rate of 85.7% in both Fig. 1.7.2 (Fall) and 1.8.2 (Spring). In Fig. 1.7.2 Asians have the second highest A/B rate of 48.1%, much lower rate than 85.7%. Fig. 1.8.2 shows Whites with the second highest A/B rate of 28.1%, an even more significant drop from 85.7%. Both graphs show Blacks with the lowest A/B rate.

Drop/Fail Rate

Besides the average final exam score and A/B rate, the drop/fail rate is an important measure of student performance. It represents failure to complete the course with a passing grade, and deserves particular attention so this problem can be alleviated. Fig. 1.1.3 shows
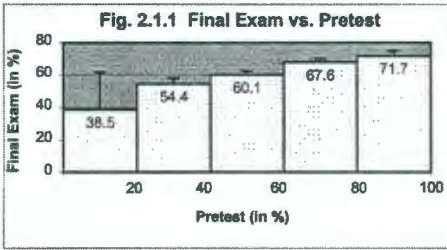
that in Fall, as pretest scores increase, drop/fail rates decrease. Drop/fail rate is a low 4.5% when pretest scores are 60-79% and no students dropped or failed when they scored 80% or better on the pretest. However, these two categories consist of only 16% or all the students. Requiring 60% on the pretest as an entrance criterion would reduce drop/fail rates, but would dramatically reduce class sizes. Fig. 1.2.3 shows a similar relationship between drop/fail and pretest scores in Spring. If 60% were the pretest cut-off in Spring, only 35% of students would have met the requirement.

One would think that attending class more often would yield a better chance of success. Fig. 1.3.3 and 1.4.3 validate this belief, as we see that fewer students dropped or failed when percent attendance increased.

Looking at drop/fail versus gender, we see no gender differences in Fall's Fig. 1.5.3. On the other hand, Spring's Fig. 1.6.3 shows that males dropped/failed an average of 11.1% more than females. This result is consistent with the slightly better female performance on the final exam and slightly higher A/B rates discussed earlier.

Drop/fail rates remained steady regardless of a student's ethnic group, with one exception—non-residents. Fig. 1.7.3 shows the drop/fail rate in the 30% range for Blacks, Whites, Hispanics, and Asians. Non-residents had the lowest rate of 14.3%. Fig. 1.8.3 is similar in that Blacks, Whites, and Hispanics averaged in the high 50% range. Asians had a 33.3% rate and non-residents had the lowest rate of 14.3%.
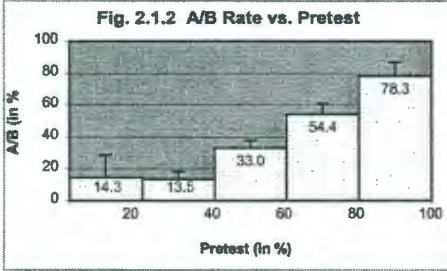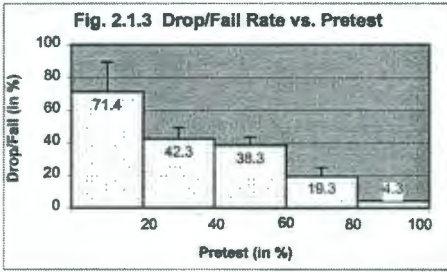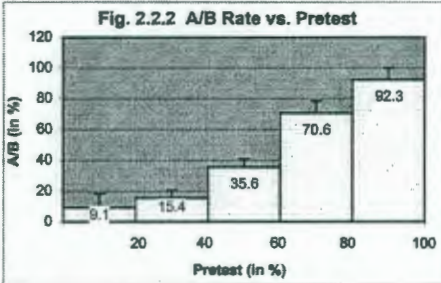
Figures 2.1.1 – 2.1.3
Pretest Scores versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Calculus I (Fall 1999)

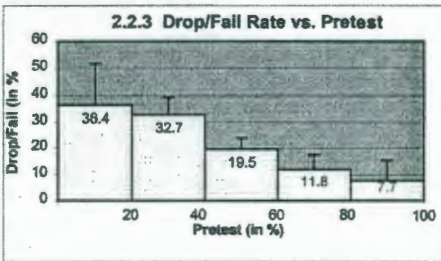Fig. 2.2.1 Final Exam vs. Pretest
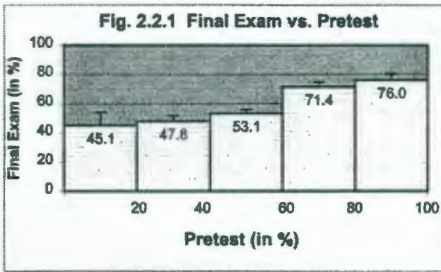
Pretest(%)	N	%
< 20	7	5
20 - 39	40	26
40 - 59	67	44
60 - 79	27	18
> 79	13	8

Fig. 2.2.1   Corr: .39

Fig. 2.2.2 A/B Rate vs. Pretest

Pretest(%)	N	%
< 20	11	6
20 - 39	52	26
40 - 59	87	44
60 - 79	34	17
> 79	13	7

Fig. 2.2.2   Corr: .47

2.2.3 Drop/Fail Rate vs. Pretest

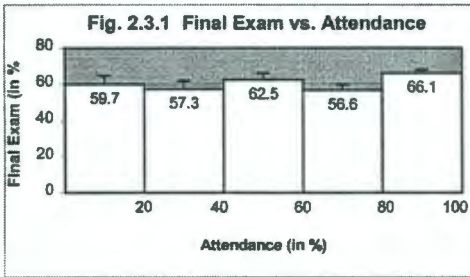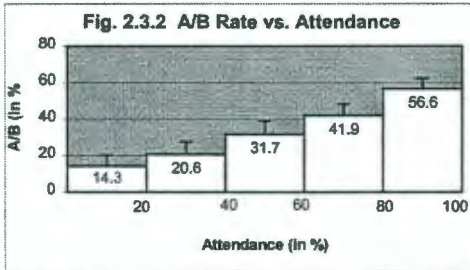| Pretest(%)	N	%
< 20	11	6
20 - 39	52	26
40 - 59	87	44
60 - 79	34	17
> 79	13	7

Fig. 2.2.3   Corr: -.20

Figures 2.2.1 – 2.2.3
Pretest Scores versus Average Final Exam Score, A/B Rate, and
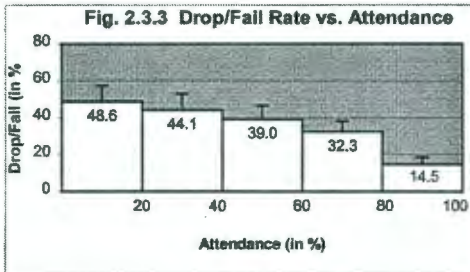Drop/Fail Rate in Calculus I (Spring 2000)

Fig. 2.3.1 Final Exam vs. Attendance

Attn(%)	N	%
< 20	17	9
20 - 39	21	11
40 - 59	25	14
60 - 79	51	28
> 79	70	38
Fig. 2.3.1	Corr:	0.12

Fig. 2.3.2 A/B Rate vs. Attendance

Attn(%)	N	%
< 20	35	14
20 - 39	34	14
40 - 59	41	17
60 - 79	62	25
> 79	76	31
Fig. 2.3.2	Corr:	0.31

Fig. 2.3.3 Drop/Fail Rate vs. Attendance

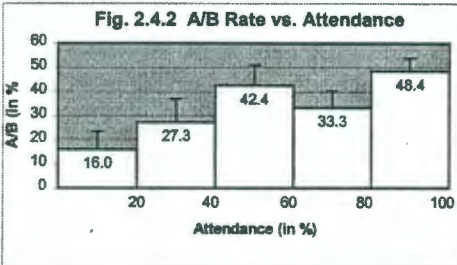| Attn(%)	N	%
< 20	35	14
20 - 39	34	14
40 - 59	41	17
60 - 79	62	25
> 79	76	31
Fig. 2.3.3	Corr:	-0.26

Figures 2.3.1 – 2.3.3
Attendance versus Average Final Exam Score, A/B Rate, and
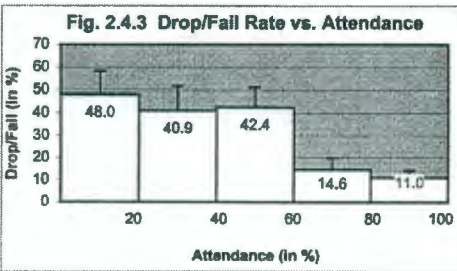Drop/Fail Rate in Calculus I (Fall 1999)

Fig. 2.4.1 Final Exam vs. Attendance
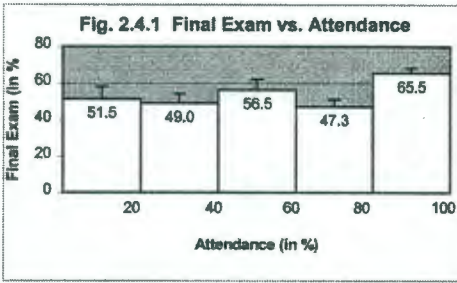
Attn(%)	N	%
< 20	13	8
20 - 39	16	10
40 - 59	20	12
60 - 79	39	23
> 79	79	47

Fig. 2.4.1 Corr: .22

Fig. 2.4.2 A/B Rate vs. Attendance

% Attn	N	%
< 20	25	11
20 - 39	22	10
40 - 59	33	15
60 - 79	48	22
> 79	91	42

Fig. 2.4.2 Corr: .20

Fig. 2.4.3 Drop/Fail Rate vs. Attendance

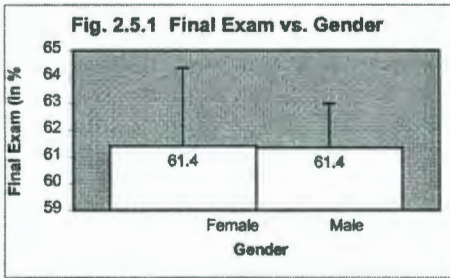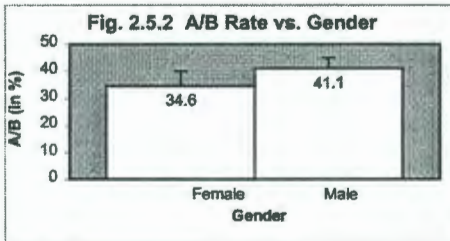| % Attn	N	%
< 20	25	11
20 - 39	22	10
40 - 59	33	15
60 - 79	48	22
> 79	91	42

Fig. 2.4.3 Corr: -.34

Figures 2.4.1 – 2.4.3
Attendance versus Average Final Exam Score, A/B Rate, and Drop/Fail Rate in Calculus I (Spring 2000)

Fig. 2.5.1 Final Exam vs. Gender

Gender	N	%
Female	59	32
Male	125	68

Fig. 2.5.1

Female: 61.4   Male: 61.4

Fig. 2.5.2 A/B Rate vs. Gender

Gender	N	%
Female	78	32
Male	163	68

Fig. 2.5.2

Female: 34.6   Male: 41.1

Fig. 2.5.3 Drop/Fail Rate vs. Gender

Gender	N	%
Female	78	32
Male	163	68

Fig. 2.5.3

Female: 33.3   Male: 32.5

Figures 2.5.1 – 2.5.3
Gender versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Calculus I (Fall 1999)

Fig. 2.6.1 Final Exam vs. Gender

Gender	N	%
Female	55	34
Male	109	66
Fig. 2.6.1		

Fig. 2.6.2 A/B Rate vs. Gender

Gender	N	%
Female	77	36
Male	137	64
Fig. 2.6.2		
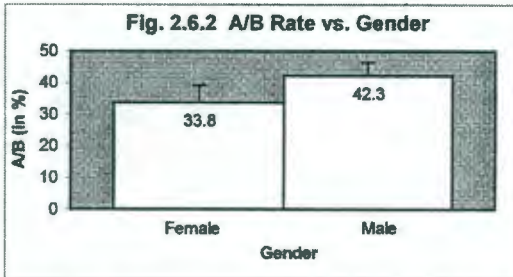
Fig. 2.6.3 Drop/Fail Rate vs. Gender

Gender	N	%
Female	77	36
Male	137	64
Fig. 2.6.3		

Figures 2.6.1 – 2.6.3
Gender versus Average Final Exam Score, A/B Rate, and
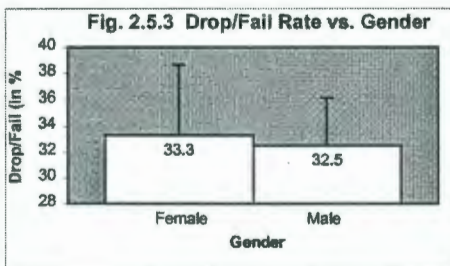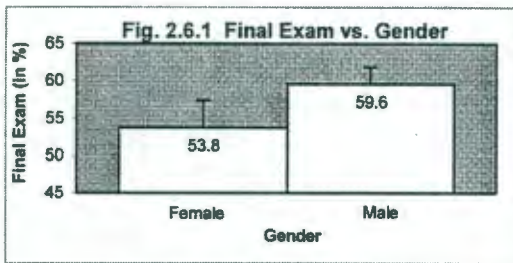Drop/Fail Rate in Calculus I (Spring 2000)
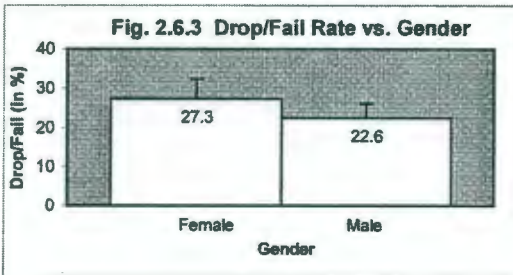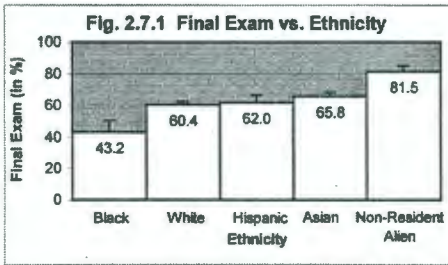
ETHNIC	N	%
Black	13	7
White	101	56
Hispanic	18	10
Asian	38	21
Non-Res.	10	6
Fig. 2.7.1		

Ethnic	N	%
Black	17	7
White	139	59
Hispanic	23	10
Asian	43	18
Non-Res.	12	5
Fig. 2.7.2		

Ethnic	N	%
Black	17	7
White	139	59
Hispanic	23	10
Asian	43	18
Non-Res.	12	5
Fig. 2.7.3		

Figures 2.7.1 – 2.7.3
Ethnicity versus Average Final Exam Score, A/B Rate, and
Drop/Fail Rate in Calculus I (Fall 1999)

Fig. 2.8.1 Final Exam vs. Ethnicity
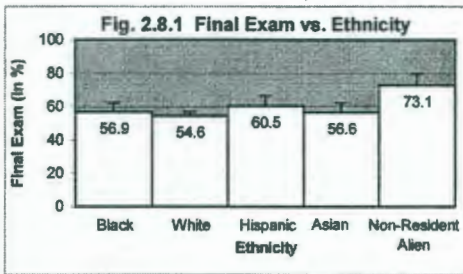
Ethnic	N	%
Black	15	9
White	92	57
Hispanic	15	9
Asian	22	14
Non-Res.	18	11
Fig. 2.8.1		

Fig. 2.8.2 A/B Rate vs. Ethnicity

Ethnic	N	%
Black	20	9
White	121	57
Hispanic	23	11
Asian	30	14
Non-Res.	18	8
Fig. 2.8.2		

Fig. 2.8.3 Drop/Fail Rate vs. Ethnicity

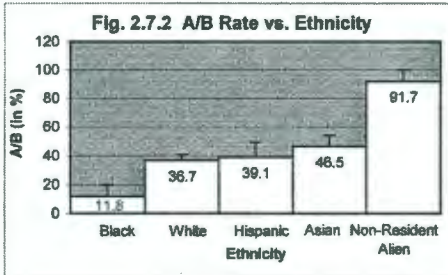| Ethnic	N	%
Black	20	9
White	121	57
Hispanic	23	11
Asian	30	14
Non-Res.	18	8
Fig. 2.8.3		

Figures 2.8.1 – 2.8.3
Ethnicity versus Average Final Exam Score, A/B Rate, and
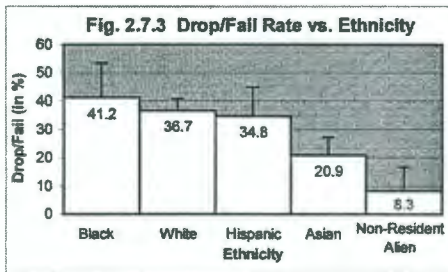Drop/Fail Rate in Calculus I (Spring 2000)

CALCULUS I DISCUSSION

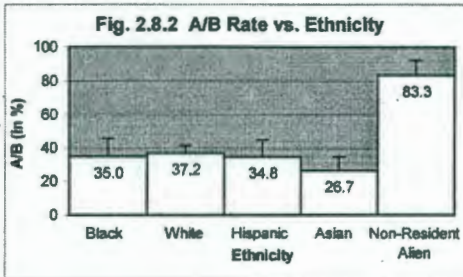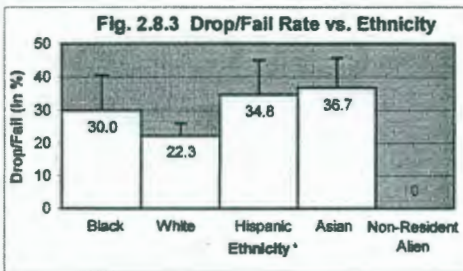The tables next to Fig. 2.1.1 and 2.2.1 tell us that an average of only 34% of students in the Fall and in the Spring scored at least 60% on the pretest; that is, about two thirds of the students enrolled in Calculus I without sufficient background knowledge to do well in the class.

Perhaps the students realized their own weakness in the beginning and attended class regularly to make up for the lack of knowledge. The table next to Fig.2.3.2 shows that 56% of the Fall students attended class at least 60% of the time. The table for Fig. 2.4.2 indicates 54% of the Spring students attending class at least 60% of the time. A smaller percentage of students had decent attendance records of at least 80%.

Comparing the pretest correlations in Fig. 2.1 and 2.2 with the corresponding attendance correlations in Fig. 2.3 and 2.4, we see that the pretest correlations are stronger than the attendance correlations, so pretest is a better predictor of performance.

Final Exam

Fig. 2.1.1 (Fall) and 2.2.1 (Spring) show a strong relationship between pretest scores and final exam scores. In both semesters the average final exam score increases steadily as pretest category increases.

Unlike pretest scores, attendance is not a strong predictor of final exam scores. Looking at Fig. 2.3.1 and 2.4.1, the > 79% attendance category has the highest average final exam, but there is no consistent increase as attendance percentage increases.

Comparing final exam to gender, we see in Fig. 2.5.1 that, on average, males and females score the same. Fig. 2.6.1 shows that males scored an average of 5.8 points higher than females, but the difference is not statistically significant.

The relationship between final exam and ethnicity is depicted in Fig. 2.7.1 and 2.8.1. From both graphs we see again that the non-resident aliens have the highest final exam average, scoring about 15% higher than the closest competitor.

A/B Rate

A/B rate has a very strong relationship with pretest scores. Fig. 2.1.2 and 2.2.2 show A/B rate increasing dramatically as pretest scores increase. The A/B rate is as high as 92.3% in Spring for those who scored at least 80% on the pretest.

Attendance relates to A/B rate just as pretest scores do. Fig. 2.3.2 and 2.4.2 show that when attendance increases, do does A/B rate.

Regarding the relationship between A/B rate and gender, Fig. 2.5.2 and 2.6.2 show that males outperform females, but the difference is not statistically significant. In Fall's Fig. 2.5.2 males average 6.5% higher than females and in Spring's 2.6.2 they average 8.5% higher. Also, in both semesters the tables indicate that males represent well over 60% of the students.

A/B rate varies across ethnic groups just like final exam scores do. Again, non-residential aliens excel with an A/B rate of 91.7% in Fig. 2.7.2 (Fall) and 83.3% in Fig. 2.8.2 (Spring). In Fig. 2.7.2 Asians have the second highest A/B rate of 46.5%, and Blacks' 11.8% is the lowest. Fig. 2.8.2 shows Whites placing second with 37.2% and Asians having the lowest A/B rate of 26.7%.

Drop/Fail Rate

Drop/fail rate has a very strong negative relationship with pretest scores. That is, dramatically fewer students dropped/failed when the pretest scores increased. In Fall's Fig. 2.1.3, the highest drop/fail rate is 71.4% for those who scored below 20% on the pretest, whereas those who scored over 79% on the pretest had a 4.3% drop/fail rate. In Spring's Fig. 2.2.3 the highest rate was 36.4% (below 20% on pretest) and the lowest rate was 7.7% (below 79% on pretest) not as broad a range as in the Fall. Requiring 60% on the pretest for entrance into this class, as a means to lower the drop/fail rate, would eliminate 66% of the students in Fall and 76% in Spring.

Like drop/fail versus pretest, the drop/fail versus attendance follows a similar pattern, as one might expect. When attendance increases, the drop/fail rate decreases in Fig. 2.3.3 and 2.4.3.

There are no gender differences in drop/fail rates. Fig. 2.5.3 shows that females drop/fail 0.8% more than males in Fall. Fig. 2.6.3 shows females dropped/failed 4.7% more in

Spring. These gender differences are not statistically significant.

Regarding drop/fail versus ethnicity, non-residents once again show the lowest drop/fail rate in Fig. 2.7.3 and Fig. 2.8.3. In Fall's Fig. 2.7.3, non-residents have a low drop/fail rate of 8.3% and blacks have the highest drop/fail rate of 41.2%. Spring's Fig. 2.8.3 shows that no non-residents dropped or failed while Asians lead with the highest drop/fail rate of 36.7%.

CONCLUSIONS

Pretest Scores

Students in Analytic Geometry and Calculus I generally demonstrate very weak prerequisite knowledge on first day pretests. However, the results show that as pretest scores improve, exam score and A/B rate tend to be higher, while the drop/fail rate tends to be lower.

Percent Attendance

The results indicate that, in both classes, attending class more often can lead to better success. Attendance is not a good predictor of average final exam score, but is a good predictor of A/B rate and drop/fail rate. However, one must be wary of this factor because attendance was not perfectly recorded.

Gender

Some slight gender differences are observed but are neither statistically significant nor consistent across the courses. In Analytic Geometry, one sees that females tend to make more A's and B's and dropped or failed less than males. The Calculus I results show the exact opposite: that males performed better.

Ethnicity

Ethnicity does not predict student performance unless a student is in the non-residential aliens group. This group of students excelled on the final exam, had the highest A/B rates, and experienced the lowest drop/fail rates. Blacks, Whites, Hispanics, and Asians did not have consistent performance records.

REFERENCES

Burton, L. (1990), *Gender and mathematics: An international perspective* (Cassell, London).

Clark, C. S. (1994), Education and gender, *The CQ Researcher*, 483-496.

Kirschner, V. (1981), Females and mathematics, *National Institute of Education Educational Resources Information Center.* Fact sheet number 1.

Sommers, C. H. (2000), The war against boys, *The Atlantic Monthly*, 59-74.

The Effects of Childcare Issues on Modern Employment and Childbearing Decisions of Women in the United States

Rachel Hock

Mentor: Dr. Joan R. Rycraft, School of Social Work

ABSTRACT

One of the major problems that working mothers face is locating high-quality, affordable daycare programs which will allow them to monitor their children's progress and development while they are at work. The purpose of the present study was to define and examine the various issues in the area of childcare which have a direct effect on the decisions that women make in terms of childbearing and employment. The research study was conducted through the use and statistical analysis of a survey instrument which was given to a convenience sample of 32 working women and 32 stay-at-home mothers who were taken from a population of female professionals, full-time workers, and students at the University of Texas at Arlington, as well as a group of stay-at-home mothers who live in the Dallas/Fort Worth area. The data revealed that more employed women would choose to have children or more children and that more stay-at-home mothers would choose to seek employment outside of their homes if high-quality, affordable childcare alternatives were made available to them.

PROBLEM STATEMENT

In this fast-paced modern society in which we live, greater numbers of women are entering the job market than at any other time in United States history for two primary reasons. These women either choose to exercise their creative and competitive natures in a variety of vocational venues or must enter the work force in order to support themselves and their families. Many women who enter the labor force are either opting to have fewer children or to remain childless. Furthermore, many stay-at-home mothers either limit themselves to participating only in part-time employment or forego involvement in paid employment outside of the home altogether. Women have to make choices and tradeoffs in the areas of employment and childbearing in order to conform to the rather rigid set of standards that American society has placed on its labor force:

> The traditional American work ethic pressures the adherence to a workday consisting of eight hours and a workweek of forty hours. The career ladder rewards those who go beyond this minimum expectation or whatever floor in terms of hours is set for a particular occupation. Depending on the profession, economic and social prestige often are gained by those who excel in their accumulation of workhours. This pressure to put in more and more time at work, away from the family, is detrimental to the personal lives of many and furthers the stratification of society (Klein, 1992, p. 206).

For women who attempt to balance the multitude of demands and responsibilities of embarking on a career while also rearing children, there is an infinite number of problems to solve and challenges to overcome. One of the major problems, which plagues stay-at-home mothers who wish to work outside of their homes, as well as working women who either have children or would like to have children, is locating high-quality, affordable childcare programs which will allow them to monitor their children's progress and development while they are at work.

LITERATURE REVIEW

In an effort to gather accurate statistical data concerning working mothers, stay-at-home mothers, and childcare alternatives, several statistical resources as well as journal

articles and studies were utilized, including "Early Childhood Program Distinctions" and "Where's Papa! Fathers' Role in Child Care." In addition, the research was well-rounded by examining books, such as <u>Kidding Ourselves: Breadwinning, Babies, and Bargaining Power</u> and <u>The Debate over Child Care 1969-1990: A Sociohistorical Analysis</u>, which concern the multitude of factors that define how and why women choose specific childcare options. Government resources from the World Wide Web were utilized in order to gather statistical information on topics such as women's educational levels in relation to birthrates and national birthrates. Web sites, which address various issues and concerns relating to stay-at-home mothers, were beneficial in the search for information about stay-at-home mothers. Finally, a recent book by Sarah Blaffer Hrdy, which is entitled <u>Mother Nature: A History of Mothers, Infants, and Natural Selection</u>, has proven to be instrumental for understanding the underlying concepts and theories behind the reproductive choices that women make throughout their lives.

HISTORICAL PERSPECTIVE

From ancient times until about a century and a half ago, women's work and family lives were inseparable due to the fact that the majority of women did not participate in paid employment outside of their own homes. Many of these women had numerous children who, because of the lack of modern conveniences and the high extent of infant dependency due to breastfeeding, took huge amounts of their time:

> Because people didn't have good alternatives to breastfeeding in those days, children kept women very busy. A woman who had seven children and breastfed them all spent up to 63 months being pregnant and between 35 and 168 months breastfeeding - that is, between eight and nineteen years of pregnancy and nursing combined (Mahony, 1995, p. 9).

These women could not hope to be anything more than housewives and mothers due to the rigid restrictions that were imposed upon them by a patriarchal society. In this type of society, such restrictions were perpetuated not only by the men within society but also by the women, who insisted that their sisters, friends, and daughters adhere to the same uncompromising standards. Even though a few women worked outside of the home either out of necessity or by choice, society still held fast to its belief that a woman's rightful place was in the home where she could be a "good" wife to her husband and provide a nurturing environment for her children. "Married women's work was also discouraged because 'the home was the heart of the nation and there could be no home life for a child where the mother went to the factory'" (Rowan, 1982, p. 75).

Before the Industrial Revolution and the rapid decrease of family-owned farms and businesses, most women still lived in close proximity to their extended and immediate families, which gave women social support networks that they could depend upon in times of need. During these times, women could depend on other women in their families to assist them in caring for their children whenever they were sick or simply needed to take short reprieves from the commonplace duties of managing households and taking care of children. From the beginning of the Industrial Revolution until quite recently, greater numbers of people began moving far away from their families into crowded cities where they lived in small apartments which forced them to lead semi-isolated existences that were virtually devoid of any social support networks. The views of the politicians and people in general

concerning women's role in society remained virtually unchanged during this period. "A modern industrialized society, it was proposed, was best served by a sex-based division of labor. This division of labor assigned the role of the economic provider to the husband, who was to see to the instrumental needs of the family, such as food and shelter, while the non-working housewife functioned in an expressive role oriented to the emotional needs of the family members" (Geerken & Gove, 1983, p. 11).

It was not until quite recently that the structure of society radically changed with the start of the sexual revolution, which broadened the roles that women played in society. Thus, women continued in the arduous tasks of managing their homes and rearing their children but without the all-important social support networks which they had once counted upon to help them in their times of need:

> It is common knowledge that a contributing factor to the child care supply problem is the inaccessibility of the traditional back-up support for child care arrangements: grandmothers and other relatives. Grandparents and other mature relatives are no longer assumed to be available to provide care for their grandchildren for two reasons: they often are employed themselves or live in a different city (Klein, 1992, p. 67).

Women suddenly gained the freedom and opportunity to take control of their destinies that they had once only dreamed of possessing, and through much hard work and dedication, they forged new frontiers within educational institutions, political arenas, the labor force, and even the armed forces. Yet, with the newly-found freedom which women now enjoyed and the technological and scientific advancements which made life easier came unexpected challenges which made life choices much more complicated:

> However, as maternal employment patterns have changed over the years, so, too, have other aspects of society and particularly the family. Family size in America has decreased, the management of a household has become more efficient, marital stability has declined, notions of what a person should expect from life have changed, women's roles have been reconceptualized, child- rearing orientations are different, and the adult roles for which children are being socialized are not the same as they once were. The selective factors that determine which women will seek employment and which will not have been altered; what was once the deviant choice is now the modal choice. It is not only the employed mother today who must justify her role and cope with possible guilt and anxiety about how it affects her children, but also the full-time homemaker who feels a need to explain her decisions and to defend her failure to contribute economically to the family and to conform to the new image of women (Hoffman & Youngblade, 1999, p. 4).

The Rise of Child Daycare Centers

It became increasingly clear that society was becoming much more open to women fulfilling their professional and personal goals. However, unlike in the past, mothers were confronting the challenges of having to separate home life from work life, whereas work and family obligations had formerly been one and the same. As more and more women entered the labor force, the need for alternative childcare arrangements increased significantly. "For mothers with children under the age of eighteen, the United States has gone from less than 30 percent in the labor force in 1960 to less than 30 percent not in the labor force today" (Hoffman & Youngblade, 1999, p. xi). To attempt to fulfill this unprecedented

need, child daycare centers began to appear in industrialized nations, including the United States. As a result, the number of children being cared for in daycare centers in the United States quadrupled between the years 1976 and 1990. In addition, during this same period, the number of child daycare centers increased threefold to an estimated 80,000 childcare facilities (not including regulated and non-regulated family-owned daycare facilities) which served between 4 and 5.1 million children (Galinsky & Friedman, 1993, p. 39).

Childlessness in the Past and Present

Although modern advancements in the areas of science and industry have tremendously improved the standard of living and have increased life expectancy, the chances that a modern-day woman will die without children are just a little lower than the chances of a Pleistocene woman, who lived between 1.6 million and ten thousand years ago, dying childless. According to a 1992 study by Debra Judge and Sarah Blaffer Hrdy, perhaps over fifty percent of women who lived during the Pleistocene Epoch died childless, whereas forty percent of all women who resided in the Sacramento Valley of California, and died between 1890 and 1984, did not have any surviving offspring (Hrdy, 1999, p. 7). Of course, the reasons behind ancient women's childlessness and modern women's childlessness are quite different. Almost all Pleistocene women of childbearing age became pregnant, but because of poor nutrition, uncontrollable environmental conditions, and incurable diseases, most pregnancies ended unsuccessfully. Even if an infant managed to be born without defect, its chances of surviving the first year of life were quite slim, and even slimmer still were its chances of surviving to adulthood.

However, in the modern-day United States, almost every infant who is born will survive; only six infants out of every hundred will fail to do so (Hrdy, 1999, p. 7). One would logically conclude that the percentage of childless women would have decreased significantly; however, this trend has not occurred. One reason that the number of childless women has not drastically decreased during the modern era compared to ancient times could be the fact that women in the professional world are strongly discouraged from starting families due to the potentially negative impact that it would have on their incomes. In contrast, men are expected to start families and often rewarded for doing so:

> Nonetheless, it is clear that getting married and having children greatly reduce the average woman's earnings. Moreover, with each child she has, her earnings drop even lower. . . In contrast, men's income rises with children. Married fathers living with their children work longer hours and seek promotions. There is also evidence that bosses give new fathers raises (Mahony, 1995, p. 13-14).

Another reason that women are opting to either have fewer children or remain childless is because more and more women are obtaining higher levels of education. A link exists between women's educational levels and the number of children to which they give birth:

> A woman's educational level is the best predictor of how many children she will have, according to a new study from the National Center for Health Statistics . . . The study, based on an analysis of 1994 birth certificates, found a direct relationship between years of education and birth rates, with the highest birth rates among women with the lowest educational attainment (http://www.cdc.gov/nchs/releases/97facts/ 97sheets/edu2birt.htm).

In addition, women who have attained college-level educations are more likely than women with less education to delay starting their families, as the following source indi-

cates, "For college-educated women, low first birth rates for women in their twenties and high first birth rates for women in their thirties point to the continuing trend of delayed childbearing" (http://www.cdc.gov/nchs/products/pubs/pubd/mvsr/supp/46-45/mv45_10s.htm).

CHILDCARE IN MODERN TIMES

One reason that the percentage of childless women has failed to decrease significantly is that modern women are being forced to choose between becoming stay-at-home mothers who are solely dependent on their partners for financial support, opting to become professionals who choose to be childless in order to advance their careers, or balancing work and family obligations in a patriarchal society. In patriarchal societies, childcare does not rank as a high priority for policymakers:

> Even when the goal is the survival of the family line, maternal and paternal interests do not necessarily overlap. Patriarchal societies are those in which patrilineal interests have, over time and by whatever means, come to prevail over strictly maternal ones. The goal is to produce offspring - often many of them - of undisputed paternity, no matter what the cost to their mothers (Hrdy, 1999, p. 264).

In fact, it seems that no matter what choices a mother makes in regard to childbearing and employment practices, she is sure to be criticized by her family members, acquaintances, and society in general. For instance, if a woman decides that she wants to be a full-time stay-at-home mother, she not only has to learn to adjust to the normal stresses of such a choice but also is often viewed in a negative and stereotypical light:

> Besides the usual stress that comes with any new job, she knows that our culture does not put much value on the job of homemaker or child-care provider and actually promotes negative images of these vocations. This negative climate can make the at-home mother's task of creating a new self-image, which should be a joyous work of personal transformation, more stressful than it need be (Sanders & Bullen, 1999, p. 67).

In addition, if a woman decides to embark upon a professional career while simultaneously taking on motherhood, she faces a bombardment of disapproval and criticism from almost all facets of society:

> But instead, a new and darker image of working moms gradually gained ground. In congressional testimony, glossy magazines, news stories, and Op-Ed pieces, moms with successful careers were reviled as selfish and materialistic, putting their own ambitions ahead of their children's needs. In 1991, the hip metropolitan magazine *New York Woman*, which was usually an outspoken advocate of working women, went so far as to compare the children of professional women with victims of domestic violence. The illustration to the feature story which ran under the cover line "Trophy Kids: Children of the Rich and Busy," portrayed a doleful child, wasting away for lack of attention, while her mom worked. Career success even became pivotal in a wave of widely publicized child custody cases. Tragically, some women lost their children when judges ruled them to be "too career-oriented" (Holcomb, 1998, p. 20).

Even though society in general has finally welcomed women into various professions within the labor force, it does not want to deal with the main problem that millions of women face every single day: the problem of daycare. Although some companies and orga-

nizations have reacted to the need for affordable, quality daycare by developing on-site daycare facilities, providing flexible scheduling, and encouraging family-friendly work environments, the majority of companies have not done very much to help working mothers balance their work and family responsibilities. Women are forced to make difficult and personal decisions concerning whether or not to have children based on their career goals and the availability of adequate childcare arrangements. Increasing numbers of working women are opting either to have fewer children or no children at all in order to advance their careers or simply maintain their current career status. In 1996, the total number of births declined to 3,891,494, which was the lowest level in two decades (http://www.cdc.gov/nchs/releases/98news/98news/natal96.htm).

For those professional women who choose to have children, securing adequate childcare arrangements can be a daunting and disheartening task. In 1990 a study found that new mothers spent an average of five weeks locating daycare programs for their infants (Galinsky & Friedman, 1993, p. 39). In addition, women with children under the age of three years have a much more difficult time locating adequate daycare arrangements. This is due to the fact that there is a shortage of available openings for children under three years old. In fact, less than ten percent of childcare facilities had openings for infants. Likewise, less than 16 percent of childcare facilities had openings for one-year-old toddlers. On the same note, less than 30 percent of childcare facilities have openings for children who are two years old (Galinsky & Friedman, 1993, p. 39). Women with children who are older than two years do not have quite as hard a time locating daycare facilities for their children.

Some women are forced to compromise whenever circumstances dictate and must settle for part-time employment. Married women are more apt not to utilize child daycare programs and instead leave their children in the care of spouses when they participate in part-time employment rather than full-time employment. In fact, approximately thirty-one percent of children with mothers who were both married and employed part-time in 1991 were left in the care of their fathers. In contrast, only seventeen percent of children with mothers who were both married and employed full-time in 1991 were left in the care of their fathers (O'Connell, 1993, p. 9). In general, it seems that mothers are more apt to make the decision to stay at home whenever problems arise because of childcare arrangements:

> A study of the Illinois Aid to Families with Dependent Children (AFDC) population estimated that problems with childcare had caused 20% of AFDC mothers to quit school or a training program in the previous 12 months; another 20% were estimated to have returned to public assistance because of childcare problems. Parents on AFDC with poor-quality care are more likely to leave their jobs than parents with good-quality care (Hofferth, 1996, p. 57-58).

Employer-Sponsored Childcare and its Benefits

Employer-sponsored childcare has mutual benefits for both employers and working mothers. In one way, working mothers benefit from employer-sponsored childcare programs and services because they can be assured that their children are being properly cared for by capable childcare workers while they are at work. In addition, childcare programs which are sponsored by employers benefit working mothers by reducing the exceedingly high costs of childcare. Thus, working mothers who take part in employer-sponsored childcare programs do not have the problem of locating and maintaining stable, affordable childcare arrangements for their children. It would make sense that companies which offer employer-

sponsored childcare programs, such as on-site childcare facilities, are better able to attract and keep employees than companies that do not. Klein (1992) notes, "Specifically, the multiple and interrelated stresses endured by employees and family members because of inadequate child care arrangements would be alleviated; child care then functions as a social service" (p. 199). Because they are no longer preoccupied with childcare arrangements, working mothers can give their full attention to the duties of their jobs. Not only are working mothers who are employed by companies with employer-sponsored childcare programs better able to focus on their work due to the fact that they do not have to worry about their children's childcare, but these women are absent from work far less than other working mothers. In turn, whenever employees have incentives (such as employer-sponsored childcare benefits) to remain employed at specific workplaces and are better able to give their full concentration to their work, their productivity increases. According to Klein (1992), "Therefore, within corporate America child care can function to bolster an increase in productivity and profits. Child care serves the 'bottom line'" (p. 179).

METHODOLOGY

In order to gather as much data as possible concerning the ways in which childcare issues affect employment and childbearing decisions of modern women , an independent research study was conducted. This research study was designed to be a descriptive study exploring the influences of childcare issues on childbearing and employment decisions. The research study was conducted through the use and statistical analysis of a survey instrument which was given to a convenience sample of 32 working women and 32 stay-at-home mothers. The sample was derived from a population of female professionals, full-time workers, and students at the University of Texas at Arlington as well as a group of stay-at-home mothers who live in the Dallas/Fort Worth area. The majority of the stay-at-home mothers were taken from a population of stay-at-home mothers whose children participate in "Storytimes" at the Central Branch of the Arlington Public Library. Some of the stay-at-home mothers who participated in this research study were found at local shopping centers, while others were simply located throughout the community. The surveys were distributed and collected manually for the convenience of the participants. The survey instrument consisted of questions which were designed to gather general demographic data on the participants of this survey, data on their current childcare arrangements, and their opinions concerning childbearing and maternal employment. The data collected was entered into Microsoft Excel for statistical analysis. In addition, frequency distributions and measures of central tendency on respondent demographics provided an overall picture of the women who responded to the survey. Bivariate analysis was conducted to identify factors that reveal a relationship between demographic information and views on the issues of childcare.

FINDINGS

Of the sample of 64 employed women and stay–at-home mothers, stay-at-home mothers composed 50 percent of the participants, while the other 50 percent consisted of employed women. Of the 32 employed women in the sample, 9.4 percent were full-time students; 43.8 percent were full-time workers with children; 21.9 percent were professionals with children; and another 25 percent were professionals without children.
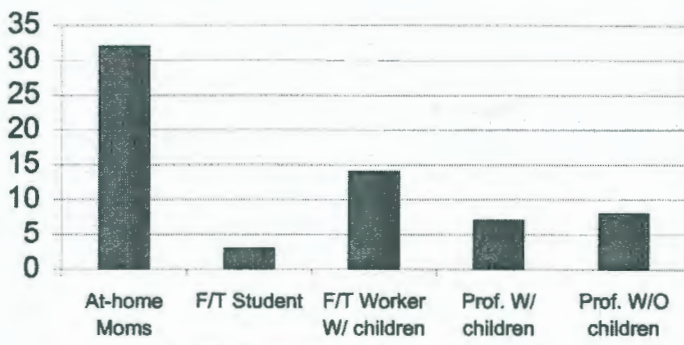
Figure 1. Participant Classifications

EMPLOYED WOMEN

Age and Marital Status

Of the participants who were employed outside of their homes, 12.5 percent were younger than 25 years old; 37.5 percent were between 25 and 35 years old; 37.5 percent were between 25 and 35 years old; 37.5 percent were between 36 and 45 years old; 12.5 percent were between 46 and 55 years old; and none were older than 55 years old. Of the survey participants who were employed, 53.1 percent were married; 15.6 percent were divorced; 12.5 percent were cohabitating; 12.5 percent had never married; 6.3 percent were separated; and none had been widowed.
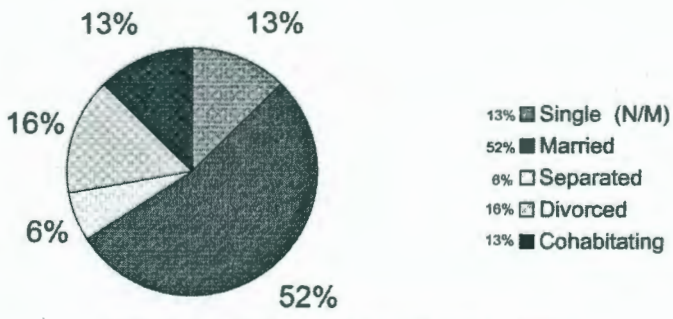

13% ☑ Single (N/M)
52% ■ Married
6% ☐ Separated
16% ☒ Divorced
13% ■ Cohabitating

Figure 2. Marital Status of Employed Women

Educational Level

In addition, 34.4 percent had graduate degrees; 18.8 percent had bachelor's degrees; 9.4 percent had associate degrees; 15.6 percent had attended college without earning degrees; 15.6 percent had either earned their high school diplomas or GEDs; only 6.3 percent had attended high school without graduating; and none held technical and trade school degrees. Among the full-time students with children, one held an associate degree; one held a bachelor's degree; and one held a graduate degree. The highest degree held by an employed woman who was not of professional status was a bachelor's degree. Only one professional with children had only a high school level education while the other professionals with children all held graduate degrees. The majority of the professionals without children held either bachelor's or graduate degrees, while only one had a high school level education.
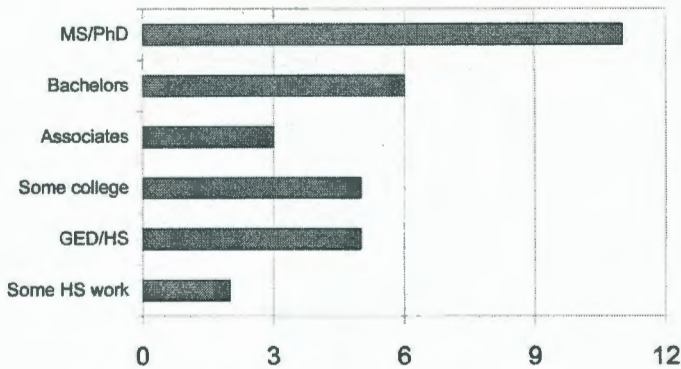
Figure 3. Educational Levels of Employed Women

Income

Only 9.4 percent reported an annual income of over $70,000; 12.5 percent reported an annual income which was between $55,001 and $70,000; 18.8 percent had an income between $40,001 and $55,000; 37.5 percent had an income between $25,001 and $40,000; 18.8 percent reported an annual income between $10,000 and $25,000; and only one participant from the sample of employed women failed to respond. One full-time student with children had an annual income between $10,000 and $25,000, while another full-time student had a yearly income between $25,001 and $40,000. The other full-time student's yearly income fell within the range of $55,001 to $70,000 per year. Most of the full-time workers with children had incomes which failed to exceed $40,000 per year. In fact, only three full-time workers had incomes greater than $55,000 per year. Two professionals with children had annual incomes between $25,001 and $40,000, while three had yearly incomes between $40,001 and $55,000. Another two professionals with children had annual incomes greater than $55,000. Two professionals without children had yearly incomes between $10,000 and $25,000, while another two had annual incomes between $25,001 and $40,000. Three professionals without children had yearly incomes between $40,001 and $55,000, while only one professional without children had an annual income greater than $70,000.

Social Support and Proximity to Relatives

A surprising 28.2 percent of the survey participants who were employed outside of their homes did not feel that they could count on their friends and relatives to provide them with social and financial support, while 69.8 percent were either sure or fairly sure that their friends and relatives would provide them with social and financial support in times of need. Another 3.1 percent of employed women had no idea if they could count on their friends and families for social and financial support. Of these same 32 employed participants, 43.8 percent resided in the same town as their families; 12.5 percent resided within fifty miles of their families; 15.6 percent resided over 50 miles away from their families; 25 percent resided in a different state than their families; and 3.1 percent resided in a different country than their families.

Household Chores

When this group of employed women was asked about the performance of the household chores, 25 percent reported that they performed all of the household chores by them-

selves; 31.3 percent reported that they performed the majority of the household chores, even though their partners performed some of the household chores; 12.5 percent stated that they equally shared the household chores with their partners; 3.1 percent reported that their partners performed the majority of the household chores, even though they performed some of the household chores themselves; none stated that their partners performed all of the household chores by themselves; and 28.1 percent stated that the question did not apply to their situations.

Children

The average number of children for employed women was 1.5. Of the 50 dependent children of the employed mothers, 10 percent were under the age of two years; 34 percent were between three and five years old; 24 percent were between six and eleven years old; 24 percent were between twelve and fifteen years old; and 8 percent were between sixteen and eighteen years old.

Current Childcare Arrangements

As for current childcare arrangements, 3.1 percent of employed women reported that they cared for their children solely by themselves; 18.8 percent reported that they shared childcare responsibilities with their partners; 18.8 percent reported that their children regularly participated in daycare programs; 6.3 percent reported that their children were regularly cared for by nannies or babysitters; 21.9 percent reported that their children were cared for by their family members; and 31.3 percent reported that the question did not apply to their situations.

Time Spent with Children

When employed respondents were asked about the amount of time that mothers should spend with their children, 9.4 percent believed that mothers should spend all of their time with their children; 28.1 percent felt that it is all right for mothers to leave their children with responsible caregivers occasionally; 62.5 percent believed that it is all right for mothers to have jobs or careers as long as their children's needs are being met by responsible caregivers while they are at work. None of the full-time students believed that mothers should spend all of their time with their children, while two students felt that it is acceptable for mothers to participate in work outside of their homes. Only one student stated that it is all right for mothers to leave their children with responsible caregivers occasionally. Only two full-time workers stated that mothers need to be with their children all of the time, while four full-time workers felt that it is acceptable for mothers to occasionally leave their children with responsible caregivers. Over half of the full-time workers believed that mothers could have jobs or careers. Of the professionals with children, only one thought that mothers should always be with their children, while the other six felt that mothers could have jobs or careers. Half of the professionals without children believed that it is all right for mothers to leave their children with caregivers occasionally, while the other half believed that it is acceptable for mothers to have jobs or careers.

Family vs. Professional Obligations

When the employed women were asked their opinions concerning family and professional obligations, 71.9 percent felt that family obligations are much more important than professional obligations; 28.1 percent believed that professional obligations are as

important as family obligations; and none thought that professional obligations are much more important than family obligations. One full-time student with children felt that professional obligations are equally as important as family obligations, while the other two full-time students believed that family obligations are more important than professional obligations. Only two full-time workers stated that family and professional obligations are equal in importance. An overwhelming twelve full-time workers felt that family obligations are more important. Two professionals with children felt that family and professional obligations are equally important, while another five believed that family obligations should take priority over professional obligations. Half of the professionals without children felt that family and professional obligations are equal in importance, while the other half of the professionals without children stated that family obligations are greater in importance.

National, State, and Local Governments' Childcare Priorities

Only 3.1 percent of the employed respondents thought that the national government places a high priority on providing quality daycare programs; 40.6 percent rated the national government's priority concerning providing quality daycare programs as average; and 56.3 percent felt that the national government places a low or very low priority on providing quality daycare programs.

None of the employed women thought that the state government places a high priority on providing quality daycare programs; 40.6 percent rated the state government's priority concerning providing quality daycare programs as average; and 59.4 percent felt that the state government places a low or very low priority on providing quality daycare programs.

Likewise, none of the employed participants thought that the local government places a high priority on providing quality daycare programs; 37.5 percent rated the local government's priority concerning providing quality daycare programs as average; and 62.5 percent felt that the local government places a low or very low priority on providing quality daycare programs.

Childbearing and Employment Decisions

When asked whether they would have children or more children if high-quality, affordable childcare options were available to them, 46.9 percent of the employed participants reported that they would either be very likely or fairly likely to have children or more children; 31.3 percent stated that they would either not be very likely or not likely at all to have children or more children; 3.1 percent stated that they were not certain; and 18.8 percent stated that the question did not apply to their situations. Two of the three full-time students stated that they would be likely to have children or more children if quality childcare options were made available to them, while the other student stated that she would not be likely to have children or more children. Half of the full-time workers with children would be likely to have children or more children, as opposed to five who said that they would not be likely to have children or more children. Of the seven professionals with children, three would have children or more children, while one would not be likely to have children or more children. The other three professionals with children stated that the question did not apply to their situations. Three professionals without children thought that they would be likely to have children or more children, while another three said that they would not be likely to have children or more children. The other two professionals without children stated that the question did not apply to their situations.
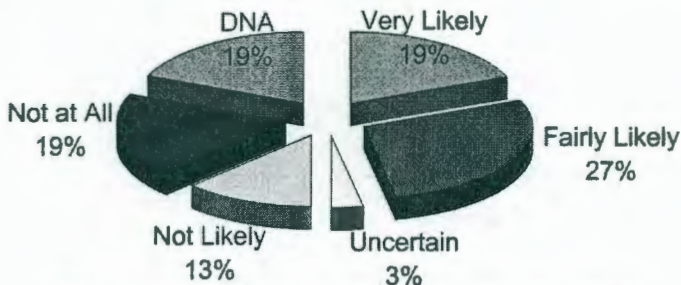
Figure 4. Potential Childbearing Among Employed Women

STAY-AT-HOME MOTHERS

Age and Marital Status

Of the 32 survey participants who were stay-at-home mothers, 9.4 percent were younger than 25 years old; 43.8 percent were between 25 and 35 years old; 37.5 percent were between 36 and 45 years old; 6.3 percent were between 46 and 55 years old; and one was older than 55 years old. Of the stay-at-home mothers who participated in the survey, 84.4 percent were married; 6.3 percent were divorced; 6.3 percent were cohabitating; 3.1 percent had never married; none were separated; and none had been widowed.

Figure 5. Marital Status of Stay-at-Home Mothers
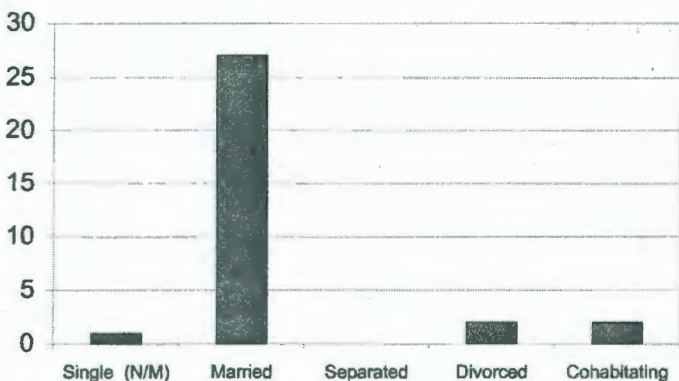
Educational Level

None of the stay-at-home mothers held graduate degrees; 31.3 percent had bachelor degrees; none held associate degrees; 31.3 percent had attended college without earning degrees; 21.9 percent had either earned their high school diplomas or GEDs; only 15.6 percent had attended high school without graduating; and none held technical and trade school degrees.
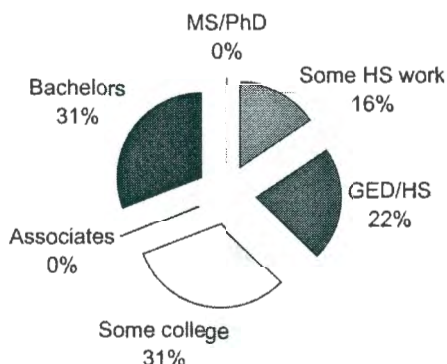
Figure 6. Educational Levels of Stay-at-Home Mothers

Income

Interestingly, 18.8 percent reported an annual income of over $70,000; 28.1 percent reported an annual income which was between $55,001 and $70,000; 25 percent had an income between $40,001 and $55,000; 9.4 percent had an income between $25,001 and $40,000; 15.6 percent reported an annual income between $10,000 and $25,000; and only one participant from the sample of stay-at-home mothers did not respond.

Social Support and Proximity to Relatives

Only 6.2 percent of the survey participants who were stay-at-home mothers did not feel that they could count on their friends and relatives to provide them with social and financial support, while 90.7 percent were either sure or fairly sure that their friends and relatives would provide them with social and financial support in times of need. Another 3.1 percent of the sample had no idea if they could count on their friends and families for social and financial support. Of these same 32 stay-at-home mothers, 34.4 percent resided in the same town as their families; 34.4 percent resided within 50 miles of their families; 15.6 percent resided over 50 miles away from their families; 15.6 percent resided in a different state than their families; and none resided in a different country than their families.

Household Chores

When this group of stay-at-home mothers was asked about the performance of the household chores, 28.1 percent reported that they performed all of the household chores by themselves; 62.5 percent reported that they performed the majority of the household chores even though their partners performed some of the household chores; 6.3 percent stated that they equally shared the household chores with their partners; none reported that their partners performed the majority of the household chores, even though they performed some of the household chores themselves; none stated that their partners performed all of the household chores by themselves; and 3.1 percent stated that the question did not apply to their situations.

Children

The average number of children for stay-at-home mothers was two. Of the sixty-five dependent children of the stay-at-home mothers, 23.1 percent were under the age of two years; 21.5 percent were between three and five years old; 35.4 percent were between six

and eleven years old; 12.3 percent were between twelve and fifteen years old; and 7.7 percent were between sixteen and eighteen years old.

Current Childcare Arrangements

As for current childcare arrangements, 53.1 percent of stay-at-home mothers reported that they cared for their children solely by themselves; 43.8 percent reported that they shared childcare responsibilities with their partners; only one reported that her children regularly participated in daycare programs or "Mothers' Day Out" programs; none reported that their children were regularly cared for by nannies or babysitters; and none reported that their children were cared for by their family members.

Time Spent with Children

When asked about the amount of time that mothers should spend with their children, 9.4 percent believed that mothers should spend all of their time with their children; 59.4 percent felt that it is all right for mothers to leave their children with responsible caregivers occasionally; 31.3 percent believed that it is all right for mothers to have jobs or careers as long as their children's needs are being met by responsible caregivers while they are at work.

Family vs. Professional Obligations

When the respondents were asked their opinions concerning family and professional obligations, 93.8 percent felt that family obligations are much more important than professional obligations; 6.3 percent believed that professional obligations are as important as family obligations; and none thought that professional obligations are more important than family obligations.

National, State, and Local Governments' Childcare Priorities

Only 6.3 percent thought that the national government places a high priority on providing quality daycare programs; 37.5 percent rated the national government's priority concerning providing quality daycare programs as average; and 56.3 percent felt that the national government places a low or very low priority on providing quality daycare programs.

Only 3.1 percent thought that the state government places a high priority on providing quality daycare programs; 40.6 percent rated the state government's priority concerning providing quality daycare programs as average; and 56.3 percent felt that the state government places a low or very low priority on providing quality daycare programs.

Only 3.1 percent of stay-at-home mothers thought that the local government places a high priority on providing quality daycare programs; 40.6 percent rated the local government's priority concerning providing quality daycare programs as average; and 56.3 percent felt that the local government places a low or very low priority on providing quality daycare programs.

Childbearing and Employment Decisions

When the stay-at-home mothers were asked whether they would choose to embark on either a job or a professional career if high-quality, affordable childcare programs were made available to them, 62.5 percent stated that they would either be very likely or fairly likely to embark on a job or professional career. Another 34.4 percent of the respondents reported that they would either be very unlikely or not likely at all to start a job or career

even if quality childcare programs were available. Only 3.1 percent were not certain whether they would start a new job or career if high-quality, affordable childcare programs were made available to them. Among stay-at-home mothers who held bachelor degrees, five would be likely to start jobs or careers, while the other five would not be likely to embark on new jobs or careers. Seven stay-at-home mothers who had been to college, but had not earned college degrees, stated that they would be likely to start jobs or careers, while only three who had been to college without earning degrees said that they would not be likely to embark on jobs or careers. Of the stay-at-home mothers who had high school level educations, five stated that they would be likely to start working if quality childcare was made available to them, while only one said that she would not be willing to start a job or career. Of the stay-at-home mothers who had not graduated from high school, three stated that they would be likely to start working, while the other two did not feel that they would be likely to work outside of their homes.
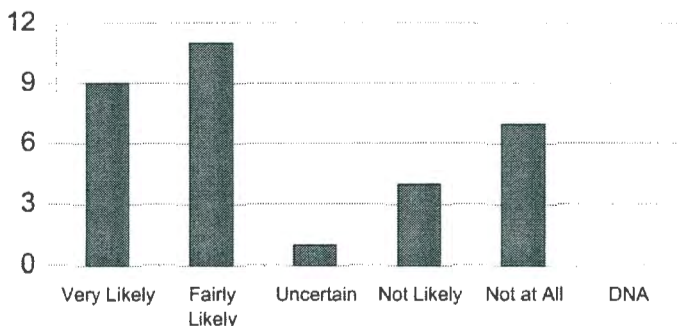

Figure 7. Employment Potential Among Stay-at-Home Mothers

DISCUSSION

The results that were found were consistent with the original hypothesis which stated that more employed women would have children or more children and that more stay-at-home mothers would seek employment if high-quality, affordable childcare options were made available to them. Although the majority of employed women stated that they would opt to have children or more children, it would seem logical that there might have been a larger majority which would opt to reproduce if there had been a greater number of employed participants in their twenties and early thirties (50 percent of the employed participants were over 35 years old). Another possible reason that there was not a larger number of employed women who were willing to have children or more children was that, as a group, these women had less social support than the group of stay-at-home mothers. It would seem logical that women with less social support would opt to have fewer children than women with adequate support networks. In addition, compared to the stay-at-home mothers, fewer employed women had families that lived in close proximity to them, and it would make sense that the employed women who are farther away from their families would be more reluctant to have children than the employed women who lived close to their relatives. A final reason for the rather low number of employed women who would have children or more children is that many of the employed participants had higher education degrees, and it is a known fact that as educational levels increase, the number of children which women have decreases.

An overwhelming majority of the stay-at-home mothers would choose to work if high-quality, affordable childcare options were made available to them. This finding makes a great deal of sense for a number of reasons. One factor which apparently plays a role in the decision of whether to work is the age of the stay-at-home mother. Interestingly, when compared to mothers older than 35 years of age, a larger percentage of stay-at-home mothers who were in their twenties and early thirties would like to work outside of their homes. Perhaps, the younger stay-at-home mothers had been reared to have a more liberal set of expectations concerning the roles of mothers in society than those of the older stay-at-home mothers.

Another reason that the majority of stay-at-home mothers in this study might choose to work could be due to the fact that the majority of these women are fairly well educated. It would make sense that the stay-at-home mothers with the most formal education would also be the ones most likely to seek employment. However, this hypothesis is not supported by the findings of this study. Although the percentages of mothers who had completed high school or had attended college without obtaining degrees were exceedingly high, the stay-at-home mothers with bachelor degrees had the lowest percentage of mothers who would seek employment outside of their homes. Possibly, the stay-at-home mothers with bachelor degrees had made conscious decisions to stay home and educate their children prior to even starting their families.

Although no strong correlation was found between the number of children that women have and their willingness to seek employment outside of their homes, it seems that women with one child might be much more likely to work outside of their homes than women with more than one child. Mothers with one child would possibly be more willing to work because the costs of childcare would be less expensive than the costs for more than one child. It is also likely that mothers with only one child would welcome the opportunity to participate in the additional activities that employment outside of their homes would offer as their children grew older. Interestingly, there was a major discrepancy between the beliefs that stay-at-home mothers held about working while rearing children and their likelihood to engage in employment outside of their homes. Only 31.3 percent believed that it is all right for mothers to work, while 62.5 percent stated that they would be willing to seek employment.

A larger percentage of stay-at-home mothers, as compared with employed women, continued to perform all or most of the household chores. This finding is not too surprising since stay-at-home mothers are presumably at home a larger percentage of the time than their partners and presumably take on more responsibilities around their homes. However, over 70 percent of all women who participated in this survey reported that they performed either all or most of the household chores. According to this finding, even though more and more women are working outside their homes, most men do not engage in the majority of household chores.

Approximately 60 percent of the employed women and the stay-at-home mothers felt that the national, state, and local governments had low to very low priorities when it came to providing childcare programs and benefits. Furthermore, approximately 40 percent of the survey participants rated as average the national, state, and local governments dedication to providing childcare programs and benefits. These findings indicate that the government is not satisfying the public's need for high-quality, affordable childcare programs.

CONCLUSION

Childcare is a determining factor in women's decisions regarding childbearing and employment. In order to provide more effective childcare alternatives, all three levels of the government should work together to subsidize high-quality, low-cost childcare programs, which would be available to everyone on a sliding scale payment basis. Another way in which to improve childcare in the United States is for the national government to set a separate minimum wage for childcare workers. A separate and higher minimum wage would possibly decrease the astoundingly high rate of childcare provider turnover. Setting a separate minimum wage for childcare providers might also attract people of a higher caliber to the childcare industry, which would be good news for everyone involved. However, until the government takes action to improve childcare alternatives, the problem with childcare will continue to exist, and, as a result, women will continue to have to make difficult decisions concerning employment and childbearing.

REFERENCES

Birth and fertility rates by educational attainment: United States, 1994. (1994). Retrieved June 21, 2000 from the World Wide Web: http://www.cdc.gov/nchs/products/pubs/pubd/mvsr/supp/46-45/mv45_10s.htm.

Galinsky & Friedman. (1993). Early childhood program distinctions. Education before school: investing in quality child care. Families and Work Institute. 39.

Geerken, M., & Gove, W. (1983). At home and at work. Beverly Hills, California: Sage Publications, Inc.

Hofferth, S. L. (1996). Child care in the United States today. Financing Child Care. 6(2).57-58.

Hoffman, L. W., & Youngblade, L. M. (1999). Mothers at work: effects on children's well-being. Cambridge, United Kingdom: Cambridge University Press.

Holcomb, B. (1998). Not guilty: the good news about working mothers. New York, New York: Scribner.

Hrdy, S. B. (1999). Mother nature: a history of mothers, infants, and natural selection. New York: Pantheon Books.

Klein, A. B. (1992). The debate over child care 1969-1990: a sociohistorical analysis. Albany, New York: State University of New York Press.

Latest birth statistics for the nation. (1996). Report of Final Natality Statistics. Retrieved June 21, 2000 from the World Wide Web: http://www.cdc.gov/nchs/releases/98news/98news/natal96.htm.

Mahony, R. (1995). Kidding ourselves: breadwinning, babies, and bargaining power. New York, New York. BasicBooks.

Mother's educational level influences birth rate. (1994). Retrieved June 21, 2000 from the World Wide Web: http://www.cdc.gov/nchs/releases/97facts/97sheets/edu2birt.htm.

O'Connell. (1993). Where's papa! fathers' role in child care. Population Reference Bureau.

Rowan, C. (1982). 'Mothers, vote labour!' the state, the labour movement and working-class mothers, 1900-1918. In R. Brunt & C. Rowan (Ed.). Feminism, culture and politics. London: Lawrence and Wishart.

Sanders, D., & Bullen, M. M. (1999). Staying home. Wauconda, Illinois: Spencers & Waters.

Domestic Violence:
How African-American Women Cope and Survive

Julie B. Williams
Mentor: Peter Lehmann DSW, School of Social Work

ABSTRACT

Domestic violence is a serious problem in the African-American community. According to West (1999, pp. 1-3), "Intimate violence represents an alarming problem seen in this culture when traced through the history of African-American women in this country. Ever since the agonizing journey by their African ancestors to the Americas as cargo on slave ships, male physical and sexual violence has been a reality in the lives of black women. The moral consequences of this intimate violence include psychosocial as well as spiritual wounds and obstacles of the wholeness and human well-being of black women." This qualitative research paper examines the coping and survival experiences of African-American women who had left their abusive partners. Culture-specific interviews were carried out in order to understand some of the effects of coping and survival methods of African-American women twelve to eighteen months after leaving the violent relationship and establishing a life for themselves.

INTRODUCTION

Domestic violence crosses ethnic, racial, age, sexual orientation, and socioeconomic lines. By the most conservative estimate, one million women each year suffer nonfatal violence by an intimate partner (American Bar Association, 1997). Domestic violence is a system of abuse by a spouse or domestic partner involving psychological control, verbal harassment and abuse, bodily injury, and/or sexual assault (City Club, 1997). An overwhelming majority of domestic violence victims in heterosexual relationships are women (90-95%), while as many as 95% of domestic violence perpetrators are male. Domestic violence is statistically consistent across racial and ethnic boundaries. Victims may experience domestic violence at any age; however, women aged 19-29 reported more violence by intimates than other age groups, whereas women aged 46 or older are least likely to be battered by an intimate. Domestic violence also cuts across all professions, and victims have included doctors, psychologists, lawyers, ministers, business executives, and housewives (American Bar Association, 1997). Sadly, social and intimate violence overlap in African-American women's lives.

In the words of West (1999, pp. 11-12), "One defining characteristic of violence against women is secrecy and unfortunately subjection to the indignities." West includes a number of ways silence is imposed on African-American women victim-survivors. These include but are not limited to: 1) not being heard or not being listened to because of stereotypes of public representations of her moral inadequacy, 2) not being permitted to speak out because of the rules of the abuser, the courts, the church, and the patriarchal culture, and 3) censorship to the point of losing one's sense of self because she is overwhelmed by the varying and specific definitions of appropriate behavior that are given to her. When women do defy such forms of tyranny, it is a noisy silence that is broken. Their defiance, therefore, provides critical public witness to African-American women's self-expression and liberation.

According to the literature, leaving an abusive relationship is not an easy task for an African-American woman. Certainly, women who leave abusive relationships almost always acknowledge that violence is a serious problem and that they deserve to be treated with respect and to feel safe from physical harm. However, the literature also acknowledges

women need to take other steps. These include: 1) go to a safe place which is unknown to the partner (e. g., shelters, a friend or relative's home, or a church) and take your children, 2) have easy access to money, 3) memorize emergency telephone numbers, 4) and get professional help, preferably by talking to someone about the problem.

Healing will begin if the victim talks about the secrets. In fact, it's the abused woman's only hope. She can join support groups that are available in the community: the Department of Family and Children's Services provides programs such as "Mother's Day Out." She can take legal action: a protective order (protection within a 500 foot radius); if violated, it is a criminal misdemeanor in which the first offense elicits a fine, the second offense involves some jail time; a third offense is a felony. Finally, she can divorce her partner and move on: the day of staying in the relationship to be a loyal, subservient martyr is over. Clearly no one deserves to be abused (Brinegar, 1999).

This paper focuses on some coping strategies African-American women used who had left abusive relationships. Following the literature review, this paper examines four culture-specific questions about coping. Finally, the results and discussion sections analyze the findings of these questions.

LITERATURE REVIEW

Part One:
Why African-American Women Stay or Leave the Relationship

A woman is abused every fifteen seconds in this country. By the most conservative estimates, each year one million women suffer nonfatal violence by an intimate partner. Nearly one in three adult women experience at least one physical assault by a partner during adulthood (American Bar Association, 1997). Furthermore, it is estimated that husbands, boyfriends, and partners routinely batter, to some degree, women of all cultures, occupations, and ages. Abuse against women is a serious problem and tends to consist of a pattern of behaviors rather than a one-time occurrence. Battering is nothing more than the establishment of power, control, and fear in a relationship through abusive behaviors (Bureau of Justice Statistics, 1994).

Most often, for abused women, the decision to stay or to leave is not made with finality at any single point in time, but instead unfolds over time, and represents the most fundamental and difficult decision women may face (Barnett & La Violette, 1993). One's ability to cope and one's level of self-esteem help the individual to adapt psychologically and socially to stressful situations by building respect within herself. Women who are employed outside the home and who perceive themselves as being able to meet their own financial needs are more likely to terminate their relationships than women who depend on their abusers for economic support (Strube & Barbour, 1983). For women, there are three key variables that aid them to make the decision to leave an abusive relationship: source of income, transportation, and a method of child-care (Gundolf, 1988).

The African-American perspective on the leaving process can be compelling. Often the decision to leave an abusive relationship strongly reflects the victim-survivor's perception of the relationship and how that relationship compares to others. Some of the catalysts provoking the decision to leave include: (1) increasing levels of violence, leading to the realization that battering could be fatal; (2) limited resources, which make a woman feel trapped and isolated; (3) a cessation or decrease in the level of remorse expressed by the

abuser for the abuse; (4) loss of hope that things will get better; (5) the growing visibility of the violence - occurs more in public; and (6) the level of concern or outrage expressed by relatives or friends (Jasinski & Kantor, 1997). Ultimately the African-American woman's inner drive, feeling of self-worth, and concern for the welfare of her children will guide her in the decision to leave.

Part Two:
How African-American Women Cope: Whether They Choose to Stay or Leave

Domestic violence is a hidden problem, personally painful, one that family members, even the victims, try to cover up (City Club, 1997). Coping is viewed as a psychological resource that helps individuals adapt psychologically and socially to stress. A strong sense of self-worth and the desire to change their circumstances help women to cope and leave their abusers (Folkman & Lazarus, 1985). African-American culture and its importance in describing how women cope include: 1) family and the role of the woman, 2) religion and spirituality, 3) centrality of music and song as expressions of self, and 4) social support (Dubrow, Nader, & Stamm, 1999).

First and foremost, family ties are vitally important. Families provide the basics of life including food, shelter, love, guidance, and identity. African-American women socialize their children to cope with the realities of being Black in a society dominated by Whites. By having such a role, women function as a buffer between their children and a racist world. They may teach children to limit their dreams to those appropriate for a lower station in life as a protective measure and to do some sort of service work. Conversely, African-American women, particularly those who have succeeded, may teach their children to take pride in themselves and their heritage and rely upon their sense of self-worth to combat the negative messages from the world and to strive for something exceptional (Dubrow, Nader, & Stamm, 1999). This socialization to the world helps African-American women realize and understand the world and how it works. When they step back and take a look at what is happening to them, it is hoped that they can break the cycle of domestic violence. They can put their feelings and desires first and then function independently.

Secondly, African-American women voice a connection to the power within themselves through the expression of a religious faith - predominantly Christianity. One of the most prevalent aspects of faith is a devotion to a personal, living God. In this context, God is everything—omnipotent. He dwells within the individual and lends His power to the individual, particularly in times of strife, including the trauma of violence. In doing so the world may seem less oppressive and overwhelming. The indignities and injustices are easier to bear, and the joys of life easier to grasp. Consequently, religious expression to women suggests that when one is utterly forsaken, left alone, betrayed, and deserted, God is still there (Dubrow, Nader, & Stamm, 1999). Churches can play a vital role in developing community norms and values opposing domestic violence. Churches and other religious organizations can provide a wide variety of support services to help victims of domestic violence. These include:

- Taking a strong stand against domestic violence, making it clear that spousal abuse is unacceptable behavior and that it is violence, that breaks the covenant of marriage, not the wife's departure.
- Providing individual counseling, rather than inappropriately forcing a woman into couples counseling with her batterer.

- Providing parenting classes, assertiveness training, and support groups for women.

African-American women feel encouraged to make the break from the domestic violence when their stories are accepted as true, thereby validating their worth as human beings. This promotes self-esteem and aids in the coping process (City Club, 1997).

A third characteristic includes the power of song and music to tell a story, thereby recreating the true emotion of the experience. Frederick Douglass exclaimed, " …[the songs] breathe a prayer and complaint of soul boiling over with bitter anguish. Every tone was a testimony against slavery, and a prayer to God for deliverance from chains…" (Dubrow, Nader, Stamm, 1999, p. 34) Songs called the blues declare the depth of feeling in the singer and musician. They cry out in agony, pain, loss, rage, and occasionally in lusty ecstasy. Music and songs are seen to claim the freedom, humanity, and vivacity of the Black soul. Rap and hip-hop music are contemporary versions of ancient traditions in which music tells stories that testify to the reality and challenges of the world. The music puts oppression on display. Currently, these songs speak to the hostility and tension between the police and Black youth and reveal the inner dynamics of the pain within Black communities. The songs challenge, brag, defend, and confront. They are the voice of the urban African-American articulated forcefully and boldly. These works provide insight into the African-American perspective on life, even while speaking the crudest and hardest language; they declare power and vitality within the African-American self (Dubrow, et al., 1999). African-American music and songs allow the listener to identify with the plight of other abused women of her race, thereby giving the listener comfort in knowing that she is not alone and that there is a way out. The realization that circumstances are not hopeless helps the individual cope and provides motivation and inner drive to break the cycle of domestic violence.

Finally, African-Americans in general yearn to build cultural communities where families are supported, children are raised, and elders enjoy the respect that they deserve. Social service and community self-help are some of the oldest traditions in African-American life. Volunteer organizations, churches, nonprofit organizations and informal helping networks, such as Black Women's Health Network; sororities and fraternities; community centers; and clubs like the Lions, Elks, and Rotaries, have been, until recent decades, some of the limited resources for the destitute. This care provides both tangible and moral support to others and reasserts the power of African-Americans to help them - creating an inner source of healing. Community support to the African-American woman can provide inner energy; thus, giving to others can be a means of healing oneself (Dubrow, Nader, & Stamm, 1999).

METHOD

In this study, a qualitative approach by use of the interview process was utilized in examining and understanding some of the effects of coping and survival methods of African-American women after leaving the violent relationship. A survey interview approach was used. In this case the survey interview typically asked respondents to answer a series of questions in a face-to-face encounter. Three African-American mothers and their children from the Women's Haven Shelter in Fort Worth, Texas agreed to be interviewed.

After reviewing the literature and developing a questionnaire, the Women's Haven Center counselors asked that the questionnaire be faxed to them so they could review it. Two employees at the shelter, one African-American woman and one Hispanic woman,

critiqued the questions. With their suggestions, a re-write was completed and the questions were approved. The potential respondents, a total of four women, were contacted by the shelter counselors to see if they would be interested in participating in a study concerning domestic violence. Three of the four women agreed to participate in the study at which time they were asked to sign a consent form. They were generally informed about the questions that they would be answering. The counselors at the shelter informed the respondents that they would be contacted by telephone by a student from the McNair Scholar's Program about setting up an appropriate time to be interviewed. The interviews were set up and conducted at each woman's apartment at her convenience, usually in the evening after dinner. Each interview lasted approximately two hours. Answers to each question were transcribed onto paper.

Circumstances of the mothers varied. All the women had two or more children, lived in apartments, and had recently left long abusive relationships where the children had been exposed to the violence. The mothers gave an account of their own experiences with the violence and their children's response to that violence as well. To be considered to take part in the study, the women had to meet certain criteria: 1) all the women had to have experienced violence, to have gone through the shelter and be living and surviving on their own means, 2) participants had to be African-American, and 3) participation was voluntary.

Instruments

Psychosocial Questionnaire. The mothers were asked a number of questions about their abusive partner, family values and beliefs, support network, religious belief, cultural self-expression, and how they made sense of their ordeal. These questions stemmed in part, from Honoring Differences: Cultural Issues in the Treatment of Trauma and Loss (Dubrow, Nader, & Stamm, 1999), a book that provides recommendations for utilizing culturally appropriate healing resources.

RESULTS

The women who took part in this study ranged in age from 25 to 35 years. A total of three women were interviewed. All of the women were Protestant, worked at semiskilled occupations, had a high school diploma, and had at least two children. Further, two of the women had at least one year of college. For a time the women had been reared in single parent households with their mothers being the provider. All the women had been in long abusive relationships lasting from 5 to 15 years. Each woman had experienced long-term verbal, emotional, and physical abuse.

Domestic violence is tolerated by society as a private family matter. This view needs to change. The community can initiate this by telling the abuser that his behavior is not acceptable and hold him accountable for his actions. Domestic violence can be prevented but prevention means a change in the way society views family violence, in the way families socialize children (away from violence and toward equal respect), and a willingness to put community resources and support behind the effort (City Club, 1997). Consequently, this study sought to understand what influences the way that African-American women cope. Four questions were considered with respect to how African-American family values and beliefs shape the role of the woman; how and if religious faith and spirituality have an influence in the lives of African-American women; if music and song allow an opportunity for self-expression; if community social support and individual support networks are help-

ful (Dubrow, Nader, & Stamm, 1999).

To answer the question how African-American family values and beliefs shape the role of the woman, the respondents unanimously replied that their sense of independence and self-worth was instilled at an early age. By observing their parents, the women learned always to be in control of their lives and accomplish their goals through hard work and sacrifice. In essence they felt that they were taught to be strong-willed African-American women dependent on no one but themselves.

Next, in response to the question, how and if religious faith and spirituality have an influence in their lives, all answered yes. They emphasized their belief in God and how their faith in Him helped to deliver them from their abusive relationships. All the respondents stated that if not for the will of God they would not be here, but instead would have become another statistic. In fact, two of the three women stated they had been fearful of being killed.

When asked if music and song allowed an opportunity for self-expression, only two of the respondents agreed, while the third respondent could not identify with modern contemporary music. The first respondent that agreed stated that she identified with the lyrics of two artists, Gerald La Vert and Snoop Dogg. She stated that when she listened to the words she could understand the story and empathize with the situation and lyrics because she had gone through it herself. The second respondent stated that she identified with church music and some hip-hop artists. She stated that church music was inspiring to listen to because it helped to instill a sense of respect and values in children. She explained that she previews the hip-hop music and all together avoids the "gangster rap" songs because she wants her children to understand the difference between right and wrong and respect all human beings. The third respondent denied that the music and songs of today allowed for an opportunity of self-expression. She stated that people live their own lives and create their own history. In doing so she did state that soft rock music is pleasurable to listen to, but that it should not be seen as something to totally rely on.

Lastly, in response to the question if community social support and individual support networks were helpful, all answered "yes." Each woman identified her children as her primary source of support, followed by the Women's Haven Center counseling, including support groups. Finally, some extended family and close friends were also seen as sources of support, following this experience. In this study the support of family and friends was limited. Two of the respondents stated that family was not supportive because they tried to push them into church counseling and not get divorced from their abusive partners. The third respondent however, stated that her cousin "Anthony" had been a wonderful support to her and her family. He had assisted with the care of her children, home, and with doing some errands when she could not. All the women now try to include their children in their decision-making process. Each woman also felt that her decisions were based on what was best for the entire family.

Upon analysis of the questionnaire, it seems that the decision to leave the abusive relationship was based on traumatic events against them and/or their children - physical abuse and neglect. The respondents stated that they try not to dwell on the past, but instead look to the future. They all realize that they were not to blame for the actions of their abuser and that they were not bad. All the respondents stated that deep down they did not want their relationships to fail. However, each realized that, unless they moved on with their lives, the cycle of abuse could continue.

DISCUSSION

The purpose of this qualitative study was to ask about the coping experiences of African-American women who had left their abusive partners. Despite these long-term abusive relationships, the respondents of this study showed amazing courage in their ability to adapt to and change their circumstances and move on with their lives. This study examined how African-American family values and beliefs shape the role of the woman; how and if religious faith and spirituality influence choices; if music and song provide a mode of self-expression; and if community social support and individual support networks are helpful as a coping mechanism (Dubrow, Nader, & Stamm, 1999).

Certain cultural principles are learned at an early age - family structure, religious ideas, ways of socializing children, world-views, and so on. African-American women cope and leave their abusive partners with the support of family, church, and friends. For African-American women, when the love is replaced with anger and hatred, they must make a choice. Backed with a strong sense of independence and self-worth these women were able to manifest enough internal strength and resolve to break the cycle of abuse and leave their abusive relationship.

The African-American family values and beliefs system, as it relates to the role of a woman, center around protecting her family. Family ties are important in that there is a sense of responsibility in ensuring that each generation is afforded the opportunity to be better than the last. Families not only provide the basics of life such as food and clothing, but also promote independence and individuality as far as personal identity is concerned. The African-American female acts as a buffer between her family and society. She must socialize her children to cope with bias and racism at an early age, and in doing so, helps them prepare for life in general.

For most African-American families, including the participants in this study, the church and the belief in the Deity are of central importance. There is a deep devotion and connection to God. He is considered a "savior" to the African-American, in that His love helps to protect those that are in need. He is omnipotent, in that there is nothing that He cannot accomplish. For African-American women who participated, religious faith and belief in God helped them bear the hardships of life because they knew and felt He would not desert them. For them, this world is only momentary and there is a better place waiting for them in the afterlife.

African-Americans use music and song as a mode of self-expression of feelings and emotions. Likewise, the African-American respondents seemed to enjoy music as well. Music and songs tell stories about freedom, humanity, and the vivacity of the soul. They help provide insight on life and declare power and vitality within the self. They help to calm the mind during stressful times, thereby allowing the individual the reasoning ability to cope with a particular situation. This was also true for the women in this study. Lyrics appear to have helped them relate to their own life, thereby giving them the power and strength to move on.

Lastly, African-American women in this study used community social support and individual support networks as a means of coping with life. Like their non-abused African-American counterparts, the primary support resources included family and friends, followed by church groups, and finally, community support groups. In the African-American context the saying "Blood is thicker than water" implies that family is very close and should

be turned to for comfort and support. The church can provide some guidance for those in need of counseling and limited support, such as food, clothing, and shelter. Support groups or community centers provide informal help, such as emergency telephone numbers and contacts. African-American women help themselves through moral support of and assistance from others.

In the future, health care professionals should receive training to recognize and effectively serve victims of domestic violence and their children. They must be aware of community resources and legal remedies (e. g., legal aid) to appropriately refer them. This would include information about legal rights, such as restraining orders, protective orders, and newer anti-stalking laws (City Club, 1997). In addition, it would be important for health professionals to know about counseling services for battered women.

The only hope for a long-term reduction in domestic violence is to prevent the pattern of violence from beginning. For the African-American culture, it means assuring girls that they are human beings of worth, and that they do not ever deserve to be mistreated. Further, it means teaching boys that they can be manly without seeking to control women. Domestic violence in the African-American community is a complex issue involving many social and psychological forces. In this society the topic of domestic violence brings with it a stigma, but that must change; the general public needs to be educated. Domestic violence is everybody's business because it affects us all.

REFERENCES

American Bar Association (1997). The commission on domestic violence: Statistics. Retrieved May 31, 2000 from the World Wide Web: http://www.abanet.org/ domviol/
stats.html.

Barnett, O. W., & La Violette (1993). It could happen to anyone: Why battered women stay. Newberry Park, CA: Sage Publications.

Brinegar, J. L., Ph.D. (1999). New beginnings marriage and family, Inc. Retrieved May 21, 2000 from the World Wide Web: http://www.negia.net/~watahela/abuse.html.

Bureau of Justice Statistics (1994). Myths and facts about domestic violence. Retrieved May 21, 2000 from the World Wide Web: http://www.famvi.com/dv_facts.htm.

City Club (1997, August 1). Domestic violence—Everybody's business. Retrieved July 26, 2000 from the World Wide Web: http://www.pdxcityclub.org/report/ domvio.htm.

Dubrow, N., Nader, K., & Stamm, B. H. (Eds.) (1999). Honoring differences: Cultural issues in the treatment of trauma and loss. Pennsylvania: Taylor & Francis.

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be process: Study of emotion and coping during three stages of college examination. Journal of Personality and Social Psychology, 48, 150-170.

Gundolf, E. W. (1988). The effect of batterer counseling on shelter outcome. Journal of Interpersonal Violence, 3, 275-289.

asinski, J. L., & Kantor, G. K. (Eds.) (1997). Out of the darkness. California: Sage Publications, Inc.

Strube, M. J., & Barbour, L. S. (1983). The decision to leave an abusive relationship: Economic dependence and psychological commitment. <u>Journal of Marriage and the Family</u>, 45, 785-793.

West, T. C. (1999). <u>Wounds of the spirit: Black women, violence, and resistance</u>. New York: New York University Press.