STATISTICS IN EDUCATION

Dr. Rajesh Sharma

Statistics in Education

Statistics in Education

Dr. Rajesh Sharma (Associate Professor)
M.Ed., Ph.D. (Education), M.A. (History), Faculty of Education,
Tantia University, Sri Ganganagar

ACADEMIC
UNIVERSITY PRESS

Statistics in Education

Dr. Rajesh Sharma

© RESERVED

This book contains information obtained from highly regarded resources. Copyright for individual articles remains with the authors as indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereinafter invented, including photocopying, microfilming and recording, or any information storage or retrieval system, without permission from the publishers.

ACADEMIC UNIVERSITY PRESS

4378/4-B, Murarilal Street, Ansari Road, Daryaganj, New Delhi-110002.
Ph. No: +91-11-23281685, 41043100, Fax: +91-11-23270680
E-mail: academicuniversitypress@gmail.com

Year of Publication 2024-25

ISBN : 978-93-6284-149-0

Printed and bound by: Global Printing Services, Delhi
10 9 8 7 6 5 4 3 2 1

Contents

Preface

Statistics plays a crucial role in education, aiding educators, policymakers, and researchers in making informed decisions and evaluating educational outcomes. One essential application of statistics in education is in the analysis of student performance data. By collecting and analyzing data on student achievement, attendance, and behaviour, educators can identify patterns, trends, and areas for improvement, enabling them to tailor instructional strategies and interventions to meet the needs of diverse learners effectively.

Moreover, statistics is instrumental in educational research, facilitating the design and implementation of studies to investigate educational phenomena and evaluate the effectiveness of educational interventions. Researchers use statistical methods to analyze data collected from surveys, experiments, and observational studies, allowing them to draw valid conclusions, test hypotheses, and contribute to the evidence base for educational practice and policy.

Furthermore, statistics plays a vital role in educational assessment and evaluation. Standardized tests, assessments, and surveys are commonly used to measure student learning outcomes, evaluate programme effectiveness, and assess educational quality. Statistical techniques such as item analysis, reliability analysis, and factor analysis are employed to ensure the validity, reliability, and fairness of assessment instruments, providing valuable insights into student performance and programme effectiveness.

Additionally, statistics is indispensable in educational planning and policy-making. By analyzing demographic data, enrollment trends, and resource allocation patterns, policymakers can identify areas of need, allocate resources effectively, and develop targeted interventions to improve educational outcomes. Statistical modeling and forecasting techniques enable policymakers to project future educational needs, anticipate challenges, and develop evidence-based policies and strategies to address them.

Moreover, statistics is essential in monitoring and evaluating the implementation of educational policies and initiatives. By collecting and analyzing data on programme implementation, fidelity, and outcomes, policymakers and educators can assess progress towards goals, identify barriers to implementation, and make adjustments as needed to ensure the success of educational reforms and initiatives.

Statistics plays a multifaceted role in education, supporting decision-making, research, assessment, planning, and policy-making processes. By providing educators, policymakers, and researchers with the tools and techniques to collect, analyze, and interpret data effectively, statistics contributes to the continuous improvement of educational practice and the advancement of educational equity and excellence.

The book on Statistics in Education provides a comprehensive overview of statistical methods and their applications in educational research, assessment, and policy-making, aiding educators and policymakers in making informed decisions to enhance educational outcomes.

–Author

1

Introduction

Currently, there is considerable research into the teaching and learning of statistics at all levels of education. In higher education, statistics education is a fast growing and important area. In the next part, we introduce our perspective on activity theory pertaining to statistics education. I summarise some of the literature on statistics education relevant to my investigation and outline some issues arising from the research. Following that I introduce some aspects of my research design and approach. While full details of my methodology and procedures, I give some background to the two studies making up the project in this part–who the participants were and the setting for each of the studies. I explain the fundamentals of the Gothenburg phenomenographic method and my adaptation of it for Study Two. I outline the ideas underlying my development of the Approaches to Learning Statistics Questionnaire. I also explain how my study fits into an emerging form of research concerned with investigating individuals engaged in socially meaningful practices. I introduce my theoretical model, drawing on Leont'ev's framework. I firstly look at other important theories informing mathematics education and how activity theory fits with them. I describe some key elements of activity theory and include some major influences on its development in the former Soviet Union. I explain the structure of activity and its analysis. I frame my project, including the principal research question, in terms of an activity theory approach. Finally I summarise the key ideas of this approach.

PERSPECTIVE ON ACTIVITY THEORY

The theory of activity, based on the work of Vygotsky was expounded mainly by A. N. Leont'ev. It was an extremely influential theory in the former

Soviet Union, and was developed and extended by important psychologists there, including Luria, Gal'perin, P. Ia. Zinchenko, and, after them, Davydov, V. P. Zinchenko, as well as many others. It emphasises the role of external, practical activity in cognition.

According to Leont'ev it is an important proposition that:

- Internal psychological activities originate from practical activity, historically accumulated as the result of the education of man based on work in society, and that in separate individuals of every new generation they are formed in the course of ontogenetic development.

That is, higher mental abilities develop in the individual from activities that are rooted in the ongoing practical and communal life by which societies organise and reorganise themselves.

The philosophical foundations for this approach are the theories of Marx, Engels and Lenin. Their concern was with the economic foundations from which development arises, both individual development and societal progress. Following their theories, Leont'ev explained that people actively develop knowledge on the basis of life experiences, which have an economic, social and political context, and are mediated by cultural tools.

The idea of mediation arises from Engel's proposition that the work done by humans is mediated by the tools they develop. Vygotsky extended the idea of physical tools as mediators to include mediation of cognitive processes by psychological tools–speech and semiotic systems.

An eminent Western analyst of Leont'ev's activity theory, Wertsch, describes the activity framework as being concerned with how abilities are developed:

- To carry out socially formulated, goal-directed actions with the help of mediating devices.

While the Russian word *deyatel 'nost'* is commonly translated into English as "activity", the ideas incorporated in the construct are, to me, better embodied in the word "engagement".

This word conveys notions of purpose and affect and includes intellectual, as well as physical, processes. All of these are bound up with the social world surrounding the individual. Activity, in ordinary English usage, usually refers to physical activity–behaviour.

Indeed, in early Russian publications, activity referred to physical labour, which was mediated by tools. Later, the notion of activity was developed to include mental actions, such as remembering and reflection, and included Vygotsky's idea of mental tools as mediators, rather than only material tools of work. Activity, in the sense of engagement, both produces thought and is a product of the individual's awareness–her reflection on her environment.

According to Leont'ev activity, with its corresponding goals, means and conditions, dialectically forms and reforms individuals and their social worlds. That is, activity both orients individuals in the world in which they live and changes that world in cycles of mutual transformations.

Activity is described by Leont'ev as a functional unit of life:

- A system with its own structure, its own internal transformations, and its own development.

Throughout this thesis, I will try to explicate Leont'ev's ideas about the configuration of an activity system as a functional and dynamic unit. In this way I try to shed light on how students' learning activities develop and how they both shape and are organised by the setting surrounding them.

The activity of learning is a process in which people grapple with new information–to make it meaningful, to solve problems and to adapt to new conditions. Rather than stressing the nature of the information received by the learner, the emphasis of my investigation is on the individual acting in her social and cultural world–what the learner does, why she takes those actions and how her actions relate to the learning arena.

Varela, Thompson and Rosch describe cognitive capacities as:

- Paths that exist only as they are laid down in walking.

Their metaphor beautifully illustrates the inseparability of personal meaning and setting; the co-emergence of thought and action. These notions are fundamental to my interpretation, application and extension of activity theory. Activity theory emerged from a culture and era very different to my own. The insights offered by this theory, in part due to these very differences, highlight the problematic nature and complexity of human learning.

THEORETICAL MODEL

OVERVIEW

In order to present a model of how learning Statistics is derived from statistical activity, it is important for me to explain both the context of the learning I am investigating and the theory by which I interpret the findings. The scene for my two investigations. The research tools I use, such as naturalistic inquiry, phenomenography and statistical analyses, provide the means for investigating local occurrences. I use these tools to explore relationships and events as they occurred for particular individuals in a setting which was bounded spatially and in time.

It is by means of the theoretical framework, that is, by my application of activity theory, that I try to provide insights that transcend the particular context. To start off, I locate frameworks which draw on Vygotsky's theory in terms of some other major theories of education, particularly those which are currently important to research in mathematics education. An outline of activity theory, as expounded by Leont'ev. I introduce my approach to understanding students learning Statistics from an activity theory perspective.

COMPARISON WITH OTHER THEORETICAL APPROACHES TO LEARNING

Western researchers are increasingly finding that a Vygotskian perspective provides a helpful lens with which to view the complexities of mathematical learning. This approach takes into account the fundamental role of social interactions and historical or cultural influences on learning. It aims to understand how language and symbols mediate meaning. Vygotsky's work on cognitive development contributed to the evolution of cognitive psychology. In this framework, not only are the actions of learners important but also their goals and how the actions are situated in a sociohistorical context. For Vygotsky and the activity theorists who followed him, the context for all activity is important, and collective external activity precedes individual internal activity.

The emphasis on the melding of the individual with society is particularly relevant to the analysis of adult learning, as adults are cognitively fully developed in the biological sense but generate new knowledge through participation in activities which are socially and culturally rooted. Hence, while Vygotsky's interest was mainly in the development of the child his explanation of learning through interpersonal interactions, and Leont'ev's interpretations and ideas, are useful for extending research on adult development.

The Vygotskian framework provided a stimulus for theories of situated cognition in which a setting is defined as the relationship between an actor and the arena in which she or he acts. Cognitive skills take shape in the course of individual participation in socially organised practices. Research into situated learning recognises the cultural dimensions of learning and practices, for example, the cultural aspects of literacy; numeracy; apprenticeship and interactive expertise or working intelligence.

Issues addressed by researchers working in this paradigm often concern the effects of cultural artefacts–the devices and technologies, such as computers or internet systems, through which we "manipulate reality". Varela et al propose that knowledge is "amplified" by technology.

To them, knowledge is seen as:

- Tangibly and inextricably linked to a technology that transforms the social practices which make that very knowledge possible.

These issues are highly pertinent to a study of statistical learning, as statistics is an artefact of Western culture. It is linked to our technology and it both moulds our thinking–the way we understand our world–and is used to produce changes in our environment. Theories of situated cognition resonate with activity theory approaches in their recognition of cognition as constituted in the social, vocational and cultural life of people. Consonant with Leont'ev's theory research in this area acknowledges the role of tools in mediating cognition. However, in theories of situated learning, consideration is given to the importance of informal learning, an area little addressed by Vygotsky or Leont'ev.

One of the strengths of the conceptual frameworks provided by Vygotsky and Leont'ev is the integration of affective and motivational dimensions into explanations of cognition. There have been many researchers in diverse areas of education, who seek to show that motivational goals and intentions give meaning to learning–they are what drives the process. For example, Ausubel, Novak and Hanesian characterise all learning as meaningful learning.

Ames and Ames argue that goals:

- Provide the mechanism for filtering perceptions and other cognitive processes.

Bandura explains the link between intentions and behaviour by means of the construct of self efficacy:

- The conviction that one can successfully execute behaviour.

Studies by Volet; Volet and Chalmers and Volet and Lawrence show the importance of goals as mediators of university students' learning. Education research increasingly recognises that the attitudes, beliefs, motivation and intentions expressed by learners are important, not only for understanding the actions of those learners, but also for the insights provided into cognitive theory. However, much of the literature in the area of affect and motivation lacks a unifying theory. In Leont'ev's theory, the role of motivation and of goals in cognitive processes is systematised by means of his three levels of analysis of activity.

In mathematics education, research into affect and related areas, such as beliefs and perceptions, is extensive. The breadth and depth of these studies reflect an awareness of the importance of charting the changing trends and many dimensions of affect in learning. For example, Mcleod reviewed the research since 1970 in the *Journal for Research in Mathematics Education* on affective issues in mathematics learning, showing that such studies are central to the goals of mathematics education. These studies explore many different aspects of the affective domain. One major area of interest to mathematics educators is that of students' attitudes to learning mathematics, particularly gender effects. Fennema summarises work using attitude scales, such as the Fennema-Sherman Scale, for measuring gender effects in affective issues and achievement in mathematics.

Her research showed major differences between males and females on a number of variables, such as confidence in learning mathematics, or perceived usefulness of mathematics. Recent work on the Fennema-Sherman Scale by Forgasz, Leder, and Gardner suggests that changes in response patterns challenge the reliance that has been placed on this important instrument. They recommend revising it, particularly the scale referring to mathematics as a male domain. This indicates that important transformations are taking place in students' attitudes to mathematics. An area of research related to attitude studies concerns students' perceptions and beliefs about mathematics. Schoenfeld pointed out that students' performances on mathematical problems were often undermined by their beliefs about mathematics, such as the belief that a mathematical problem should always yield a solution quickly.

There have also been recent studies on factors affecting students' motivation to learn mathematics showing that classroom environment, goal setting and type of task are critical elements. Research in statistics education, too, confirms the importance of students' attitudes, beliefs, interests, expectations and motivation to their learning. Affective issues are related to the ways in which students evaluate and regulate their learning–key elements in defining their activities. In this regard, a helpful distinction was made by Semenov, a Soviet psychologist, who applied activity theory to the study of thought processes. He distinguished the “intellectual plane of thought”, which pertains to the development of the content of a problem, from the “personal plane of thought”, which refers to the individual’s evaluation of his or her efforts–reflections on the meaning and success of the ongoing mental activity.

While Semenov was referring to problem solving, his distinction extends usefully to students’ appraisals of studying a statistics course. In my project I will differentiate between students’ so called “objective” or practical evaluations of Statistics–evaluations in terms of extrinsic and culturally framed factors such as jobs, or higher study, and their personal assessment of it in terms of their liking for or interest in it and feelings about learning it.

A framework for the role of values in learning mathematics was developed by Southwell. This extends McLeod’s research on the relationship between beliefs, attitudes and emotions in mathematics education. Southwell sees values as closely related to beliefs but “more complicated and encompassing”. She denotes values by a triangle consisting of the three elements of valuing: cognition, affect and volition.

She does not distinguish between values and valuing, which renders her framework problematic to me. However, the three aspects of valuing she specifies, tie in with Semenov’s notion that the plane of thought has intellectual and personal aspects. Southwell’s notion that volition is an aspect of valuing also accords with Leont’ev’s insistence that activity is purposeful. Leont’ev proposes that need always stands behind thought or action. However, unlike Southwell, Leont’ev explains how purpose, needs and goals are socioculturally framed. That is, he explicates these relations in terms of individual’s activities, rather than simply stating that such relations exist. The focus on learning as “embedded in social situations, practices and cultures” is one of the key elements differentiating activity theory from constructivism, a major theory in mathematics education. While activity theorists assert that all learning is socially constituted, constructivism emphasises that all learning is the construction of meaning by the individual. In both cases action is required on the part of the learner. Constructivist ideas are drawn initially from Piaget’s work, but constructivism has developed in various forms, such as radical constructivism or social constructivism.

Constructivist thinking has been instrumental in challenging the way educators regard learning–and therefore teaching. Piaget’s work led to the idea that a mathematics classroom could be a place where children could direct their own learning, pose their own problems and discover their own mathematics. The impact of these ideas on research into mathematics education has been extensive, so much so that Ellerton asks whether we have become too comfortable with constructivism. What do we really mean by the constructivist teacher?

Relation between Constructivism and Vygotskian Approaches

Vygotskian perspectives are often contrasted with theories of learning based on constructivism. A Piagetian approach considers learning as the result of individual constructions of the environment, namely assimilation and accommodation.

Assimilation is the process whereby the individual integrates new perceptions or situations into her or his existing individual schemes, and accommodation is the individual's effort to adjust schemes to the environment. The essence of this approach is the focus on the individual and on stages of maturation as prerequisites for the development of mental facilities.

Vygotsky recognised the genius of Piaget's work and was influenced by it, though disagreeing with Piaget's conception of:

- The role of egocentrism in the developmental relationship of language and thought.

According to Vygotsky:

- The developmental uniformities established by Piaget apply to the given milieu, under the conditions of Piaget's study. They are not laws of nature but are historically and socially determined.

Thus Vygotsky criticised Piaget for his failure:

- To take into account the importance of the social situation and milieu.

Piaget, in turn, was aware of Vygotsky's theories. Indeed, Piaget modified some of his theories in the light of Vygotsky's criticisms. However, Bodrova and Leong point out that this happened after Vygotsky's death. For this reason, the works of Vygotsky's students, more than of Vygotsky himself, have some common ground with Piaget's ideas.

Bodrova and Leong express the opinion that this:

- Has caused many psychologists to erroneously consider the Vygotskian framework as part of Piaget's constructivist tradition.

Piaget, like Vygotsky, explained cognitive development, such as mathematical development, as a result of children interacting with their environments. However, while Piaget viewed learning as subordinate to development, Vygotsky insisted that learning, by which he meant formal or school learning, directly influences development.

According to Vygotsky:

- Instruction is one of the principal sources of a schoolchild's concepts and is also a powerful force in directing their evolution; it determines the fate of his total mental development.

To Vygotsky, thinking itself, together with the cognitive structure required for thought and indivisible from the function of that thinking, is generated by internalisation of social relations.

Meaning making takes on different forms in the interpretation by constructivists and Vygotskian theorists. In the former, it is a personal contribution of the learner who is involved in the education process.

In a Vygotskian approach, meaning making is bound up with cultural values, such that:

- The qualities of thinking are actually generated by the organizational features of the social interaction.

Vygotsky explained that higher mental processes develop through interaction with others.

He expressed this as follows:

- From the very first days of the child's development his activities acquire a meaning of their own in a system of social behaviour and, being directed towards a definite purpose, are refracted through the prism of the child's environment. The path from object to child and from child to object passes through another person. This complex human structure is the product of a developmental process deeply rooted in the links between individual and social history.

Knowledge to a constructivist is an individual construction; to the activity theorist it is a collective representation. This is not to intimate that constructivists deny the importance of the societal context, nor that there is no place for the individual in the Vygotskian viewpoint. Rather it is a shift in emphasis.

Saxe suggests of the Piagetian perspective:

- Social life is related to cognitive development as an external process, and the way sociocultural life may be deeply interwoven with the character of intellectual functioning is unanalysed.

In contrast, Bauersfeld calls activity theory a prototype of collectivist perspectives. A simplified overview of this perspective, provided by Bauersfeld is that learning is enculturation into pre-existing societal structures. This suggests a ready made world and knowledge, a notion for which activity theory is criticised. This enculturation is not a passive process. The individual develops through effective participation in activities. In contemporary research the term "appropriate" is used rather than "internalise" to emphasise the active role of the learner.

Radical constructivism, developed by Von Glasersfeld, has been criticised for studying:

- Human mental functioning as if it exists in a cultural, institutional and historical vacuum.

To its critics, radical constructivism presents a view of the human being as a closed system: "self-organising, self regulating, self-contained". Communication is problematic in this approach with meanings "taken as shared" rather than shared.

There have been attempts to integrate social aspects in accounts of the construction of mathematical knowledge. These add on to radical constructivist perspectives by acknowledging the "important but secondary" place of social interactions in knowledge construction. Lerman argues that adding the "social" to constructivism leads to incoherence. It fails to account for how people understand each other, or to explain the social dimension in a personal world.

Lerman proposes that the Vygotskian approach presents a different world view to the mentalism that underlies constructivist thinking, one in which:

- It is necessary to recognize the shift from a view of the autonomous cognizing subject constructing her or his subjectivity and knowing to one of the construction of human consciousness in and through communication.

Hence, the difference in thinking between Piaget's and Vygotsky's followers does not simply lie in the latter taking greater account of social interactions, as social constructivists do, anyway. Rather it is in acknowledging the primary sense in which subjectivities and positionings are constituted and reconstituted by the social events and cultural history surrounding the individual. To activity theorists, socially and culturally linked goals and needs are seen as integral to cognition, rather than as the "interference" of subordinate issues.

Ernest proposes that any form of constructivism retains portions of the radical constructivism metaphor of an:

- Evolving and adapting, but isolated organism, a cognitive alien in an hostile environment.

Hence the individual's cognition takes place in a private domain of experience. Transference from the public or social realm to this personal domain cannot be explained. This problem is not, however, restricted to constructivism. Davydov an important student of Leont'ev, postulates that the structural difference between individual and collective activity is also an unsolved problem of activity theory.

The two poles of thought represented by constructivism and Vygotskian approaches are major current theories which have in common a rejection of the transmission view of teaching and learning. They both focus on the actions of the individual in learning. Their points of difference have led to a multiplicity of models. Current approaches include: synthesising the two approaches superseding models and emerging connectionist theories of cognition or enactivism. One example of a middle position between individualism and collectivism is an "interactionist" perspective.

It takes the position that the culture of the classroom is constituted in social interactions among teachers and students. Another strategy has been to adopt complementary perspectives. Bartolini Bussi argues that rigid adherence to the principles of constructivism or to a Vygotskian perspective does not recognise the richness and complexity of the ideas of the founders. Moreover she feels that theoretical coherence should not take precedence over real life problems. Indeed Bartolini Bussi moves that such a pragmatic view is "deeply Vygotskyan", as Vygotsky himself was concerned with pressing social and cultural issues, rather than theorising. Lerman presents a critique of this view, arguing that adding bits of one theory to another, or slipping from one to the other, ignoring the contradictions, or plastering over theoretical holes in each, do not do justice to the insights and coherence of either.

There is considerable debate between protagonists of Vygotskian approaches and those of other current theories. I believe that the usefulness of activity theory lies in its systemic approach. Activity theory posits a view of learning in which personal experiences, goals, subjective perceptions and sociohistorical factors are interwoven. Rather than focusing on separate facets of learning and context, activity theory implies a commitment to investigating the learning process as a "dynamic system of meaning" in which intellect and affect unite and through which society and the individual interact and evolve. This view of learning includes notions of growth and diversity.

SOME INFLUENCES ON THE THEORY OF ACTIVITY

In developing a Marxist psychology, Leont'ev drew on the work of many philosophers, such as Marx himself and Engels, and of psychologists, particularly Rubinshtein and Vygotsky. Leont'ev's orientation to Marxism is a critical element in his theories about human consciousness and cognition. He explained the revolution brought about by Marx in the theory of cognition in terms of the role Marx ascribed to human practice in cognition. That is, Leont'ev stressed that to Marx cognition and activity were inseparable.

Leont'ev wrote:

- In reality the philosophic discovery of Marx consists not in identifying practice with cognition but in recognising that cognition does not exist outside the life process that in its very nature is a material, practical process.

Leont'ev concurred with Vygotsky who was convinced of the Marxian concept that the "human essence is constituted by social relations". That is, awareness, thought and other higher psychological functions, arise and develop in the interactions among people and their ties to the world in which they live. In Marx' philosophy, the object of knowledge is to transform the world, not just to understand it. In Soviet education, these ideas were translated into attempts "to construct a new socialist man". Both Vygotsky and Leont'ev built their psychology on a conception of Man developed by Engels and Marx.

This conception is expressed by Mellin-Olsen as follows:

- The conception is one of man as an acting person, at one time being both determined by history and determining it, being both created by society and creating it.

It was in this context of historical and dialectical materialism that Leont'ev considered activity as an object for psychology. Leont'ev's formulation of activity owes much to his many debates with Rubinshtein. Rubinshtein, who lived from 1889 to 1960 explored the 'mind body' problem stressing that activity is not only external behaviour but is inseparably linked with consciousness.

He expressed this in the formula:

- External causes act through internal conditions.

Leont'ev disagreed with Rubinshtein in that the latter considered that practical activity was the subject of study for psychology only to the extent that such activity included internal mental processes, such as perception or thinking. Leont'ev characterised this as a one sided view, in which mental activity is seen as directing physical activity, or external activity is viewed as dependent on psychological images. In Leont'ev's view, the "circle" of mental processes opens up "to meet" the external world of objects through external practical processes or activity. For example, bouncing a ball leads to a perception of its resilience.

To Leont'ev the function of activity is "transforming this reality into the form of subjectivity". That is, activity enters psychology not as one of its elements, nor as an aspect of internal processes, but through its function–the function of linking the individual's external and internal worlds. At the early stages of its development, activity must have an external form. Activity develops through reflection and regulation. Through processes of interacting with others and the environment, mental images and thoughts take place. According to Leont'ev these "deflect, change and enrich this activity".

Vygotsky's ideas and images permeate activity theory as a melody runs through a complex orchestrated score, with the harmonies, counterpoint, new melodies and even discordant notes, being provided by his students, mainly Leont'ev and Luria. It is therefore important for me to interpret Vygotsky's notions as they apply to activity theory in general and to my application of it in particular.

Vygotsky did not himself develop the concept of activity as a theoretical construct. However, he conducted empirical investigations based on the assumption of activity, emphasising the role of speech and semiotics in causing

Fundamental changes to the nature of the activity. According to Davydov and Radzikhovskii:

- The true methodological significance of Vygotsky's work consists of the assertion that activity is the explanatory principle in psychological theory.

This remark suggests that Vygotsky adopted Marx' approach to formulating a new psychology. In the last years of his short life, however, Vygotsky moved away from strict adherence to Marxist principles–at least this was the view of the Soviet authorities–resulting in his works being banned for many years in the Soviet Union.

Some of Vygotsky's colleagues and students, including Leont'ev, went on to interpret and propagate Vygotsky's sociohistorical theory of psychology within a Marxist framework, linking individual consciousness and personality with social and practical activities.

OUTLINE OF LEONT'EV'S ACTIVITY THEORY

Wertsch, commenting on Leont'ev's exposition of the tenets of activity theory summarises three major original contributions of Leont'ev to the works of Vygotsky and other theorists. The first of these is the construct of activity itself. Secondly, Leont'ev presents levels of analysis of human activity, arguing that activity can be viewed from the perspective of the milieu or context in which it takes place, from the goal directed actions which make up the activity and from the operations which depend on the conditions under which actions take place.

Thirdly, Leont'ev replaces sign systems, which are central to Vygotsky's theories, with activity as the mediator between humans and their worlds. Thus an analysis of activity explains the dynamics of how humans relate to their physical and social environments.

Activity

Activity, as a construct, occupies a major explanatory role in the psychology that dominated the Soviet Union in the sixties and seventies.

The very process of living is described by Leont'ev as:

- The system of activities that succeed one another.

Activity refers to the functional unit of human behaviour that relates the individual to his or her social and cultural world. Leont'ev and his fellow activity theorists explain that humans understand their world and develop knowledge about it by acting purposefully in it. In turn this activity changes the world.

Leont'ev rejects "positivist" notions of activity which stress activity as purely adaptation to the external world. His main criticism of this perspective is that if psychology is limited to the concept of socialisation of individuals, the structure and transformations which link humans with their society will remain a mystery. To understand these links he posits that we must investigate activity–its structure, its specific dynamics and its various forms, both overt and cognitive.

The Structure of Activity

Leont'ev identified three levels of analysis of activity.

Further expansions were provided by Zinchenko and Gordon.

- *The first level is concerned with the global aspect of activity:* This defines activity as a unit of life mediated by reflection; a frame outlining the context in which the activity takes place, for example, play, formal education or work. This level identifies the socioculturally defined milieu in which the actions occur. At this level analysis is therefore concerned with the motivation underlying and engendering the actions, the"energising function" of the activity.
- *The second level relates to the actions which make up and realise the activity:* How the task was carried out. To Leont'ev, actions are always directed towards a goal. He did not acknowledge that we sometimes do things without a conscious reason. He postulated that there may be many intermediate or partial goals to be satisfied on the way to achieving the final aim. These intermediate goals are determined consciously and with regard to the social relations in which they are set. Leont'ev gives the example of individuals beating the bushes to scare out an animal which will be caught by other, strategically placed, members of the community. The actions of beating the bushes do not in themselves have direct relevance to satisfying hunger, the motive of the activity. It is the connection of these actions to those of others and to the main purpose of the exercise which explain the actions.

- As outlined by Leont'ev actions have an intentional aspect and an operational aspect and these are carried out to realise a conscious motive. The same motive can give rise to different goals and accordingly can produce different actions. Conversely the same actions can realise different activities. For example, a statistical procedure such as a"t - test" may be carried out by a student to get an answer, with little reflection about understanding the procedure while different goals may lead to the same actions being carried out but with the focus on the underlying rationale of the procedure.
- The means by which actions are carried out are labelled operations by Leont'ev. Operations are therefore components of actions. While actions are determined by the goals which they fulfil, operations depend on the conditions, especially the tools, which delineate the exact mechanisms for carrying out the action. Tools can be physical, or"extra-cerebral" for example, computers or calculators for carrying out statistical procedures. They can also be cognitive, for example an algorithm, followed mechanically in order to get a result. Leont'ev posits that it is the fate of operations to become automatic by their mechanisation. Operations, however, are inseparable from the actions and in turn the goals which they serve. They are under conscious control as part of the actions and, through these actions, the activity in which the individual is engaged. The issue here is that different conditions lead to different compositions of actions. Solving a linear algebraic equation, for example, may be performed on an operational plane if the student has practised the process many times but not if the student is a novice in algebra. This level of analysis is particularly pertinent to the analysis of learning Statistics where the aim of many students is to become familiar with, and so automate, processes and algorithms.

The implication of viewing activity from different levels is that these provide different vantage points for investigation. At the global level, Leont'ev distinguishes activities on the basis of motive and the object towards which they are oriented. This provides one vantage point from which to understand students' learning. Activities are comprised of actions which are determined by the goals impelling them. Hence the conscious goals or needs driving the actions provides another perspective for analysis. Actions, in turn, consist of operations where the notion of operation has to do with the routine automatic aspects of carrying out a task. This level of analysis, therefore, relates to the means or resources for carrying out the actions.

Dynamics

To Leont'ev, the problem of understanding the dynamics of activity, is one of discovering the relationships that connect the individual's behaviour with physiological functions, "the work of the brain". By this, Leont'ev is not proposing a quick course in neuro-physiology. Rather he suggests an indirect approach–investigating the functional development of the mechanisms in the brain as a product of activity. This analysis must include an understanding of both phylogenetic development, that is, development which is sociohistorical or evolutionary, and ontogenetic development–the maturing of the organism. Leont'ev posits that higher order mental functions take place as a result of "mastering tools and operations". These mental systems are instated by or generated by activity.

This means that mental images are not formed by the brain; they are the function of the brain. They are generated in the transition from the "extracerebral to the intracerebral sphere". That is, there is no landing pad in the brain awaiting the arrival of images. It is by means of our actions, including mental actions such as reflection, that we transfer images from the external world to the cognitive sphere. Activity mediates between the environments surrounding humans and our internal domains.

Leont'ev stresses that activity is characterised, not by its units which are meaningless if studied in isolation, but by the systemic connections between the units and their transformations. That is, activity is not simply a sum of actions. Rather, the connections between the parts and the goal formation determine activity. The levels of activity are mobile or dynamic; actions can become activities in their own right. For example, driving a car is an activity if the person is learning to drive in order to pass a driving test.

It is purposeful, effortful action. The same actions–such as steering, braking, accelerating, may, however, be carried out as part of another activity, for example driving to work. Here the setting is work; the partial goal of the driving actions is to arrive at work–actions such as steering or braking, are the means of achieving this aim. Further, if the driver is experienced, the actions taken in driving may be automatic. They are now operations. However, in a difficult situation, such as a driver would experience in heavy traffic and wet conditions, conscious decisions will again regulate the actions, rather than a mechanical set of operations being performed.

The example illustrates that in the course of attaining an overall goal, activity may be split up into separate, successive actions which are consciously carried out with the help of operations which, in turn, may have been formed under different circumstances. The opposite happens when the individual is no longer conscious of intermediate results–the overt actions and mental reflection merge together or are consolidated in carrying out an activity.

According to Leont'ev, investigating both these aspects:

- The breaking down of the activity and the integration of its actions and operations, can only be done by studying the links. These may be all internal, as in cognitive activities, but, more often, internal activity is implemented by external actions.

The word "activity", as used throughout this thesis, refers to this framework. That is, a student's statistical activity encompasses her actions within a specific context, her goals, and the resources available to her, whether technological or her own expertise.

STATISTICS EDUCATION AT UNIVERSITY

"Statistics" said Jane, a mathematics student:

- Really fits my logic and it feels right. Maybe the men have finally invented a good mathematical study for a change.

The comment highlights the tension between knowledge as culturally endorsed and the individual's personal appraisal of it, a tension not always as agreeably resolved as in Jane's case. Statistical literacy and appreciation, particularly an understanding of data gathering, presentation and interpretation, are important components of undergraduate and graduate education in many fields. They are also essential for modern living in technologically developed countries.

Statistics is often a compulsory unit of university courses, such as Psychology, Economics, Business or a Health Science, because it is an important tool for analysing the "uncertainties and complexities of life and society". Despite this importance, many students in statistics courses are reluctant to study it.

Indeed, Cotts maintains that it is:

- Almost unarguable that the introductory statistics course is the most widely feared course on most university campuses.

This conviction is echoed in several studies.

RESEARCH IN STATISTICS EDUCATION

There has been considerable research into statistics education in the last thirty years. This interest in the teaching and learning of statistics is not only a part of the general growth spurt in mathematics education but is also a field of research in its own right. Four international conferences, the International Conferences on the Teaching of Statistics have been devoted to grappling with issues relating to statistics education.

Much topical literature concerns the teaching and learning of statistics at all levels of education. For example, in a special edition on teaching statistics of the Journal of Educational and Behavioural Statistics, Becker reviews 530 articles and dissertations which are documented on electronic databases such as the Educational Resources Information Centre. Almost all these articles (97%) were published after 1970 and about one third have appeared since 1990.

A bibliography on the available literature published between 1987 and 1994, concerning the teaching of probability and statistics, is provided by Sahai, Khurshid and Misra. This shows that the body of work being done in this area is extensive and growing.

Truran and Truran review Australasian research into the learning and teaching of stochastics, for the period 1992-1995, which reflects the current importance of probability, statistics and combinatorics to education in my region of the world. Electronically based journals such as the Journal of Statistics Education and Web discussion groups facilitate the extensive exchange of ideas and the development and distribution of research among participants in statistics education. The quantity and scope of the published literature, as well as the ongoing discussion and research, is a strong indication of current interest in statistics education. The literature and discourse concerns many topics. One important topic concerns teaching strategies. For example, Lan writes about self monitoring and Smith discusses the use of writing assignments in teaching statistics.

The role of computers, multimedia materials and other technological aids in teaching statistics is an area of increasing prominence. Individual differences, such as gender differences in learning statistics, have also been investigated. For instance, Clark proposes that female students prefer statistical questions which have a people-orientation. Preliminary results by Forbes however, suggest that, in statistical examinations in New Zealand, the form and type of examination questions, for example, whether questions involve essay writing or require using Calculus, are at least as important as their contextual embedding in determining gender preferences and performance.

As with any other field in education, what constitutes statistical knowledge and how to enhance the quality of students' learning of it, are key problems in statistics education. A major theme that recurs in the literature on higher education in statistics pertains to teaching and assessing statistics. There is general consensus in the community of statistics educators that reforms are urgently needed in the teaching of introductory statistics courses at university. Some university educators describe teaching activities and/or methods of assessment intended to address the problems - to make statistical concepts meaningful and useful - to relate statistics to the real world.

For example, Anderson and Loynes and Pfannkuch discuss what abilities and skills are needed by practising statisticians and how to teach these. Garfield Giraud and Keeler and Steinhorst describe new ways of teaching statistics using co-operative learning activities. Garfield and Hubbard present frameworks for developing assessment instruments and procedures which are appropriate for measuring students' understanding and applications of concepts in probability and statistics.

Many university educators report on the success of their statistics courses by describing enhanced performance and/or increased student satisfaction. However, as some researchers point out, little is known about how such courses would transfer to other settings. Becker concurs with this view. In her review of the literature and other resources on teaching statistics, she focuses mainly on statistics education at the university level. She reports that this literature is largely anecdotal, with less than 30% of the reviewed articles describing the results of empirical studies. Hence, extensive information on instructional strategies and resources are available to tertiary educators in statistics, either in print or through electronic media. However, research empirically evaluating the teaching and learning of statistics is less copious. Further, and in stark contrast to the literature on mathematics education, there is a dearth of studies on the teaching and learning of statistics which are framed in terms of theories of education. The literature on teaching and learning statistics at the tertiary level tends to concentrate on the “knowledge craft” of the teachers.

A further and major area of research in statistics education concerns students' understanding of particular concepts in statistics and probability theory. This research shows that many students have difficulties with and misconceptions about statistical ideas. Konold reviewed research showing that the intuitions of adults are at odds with accepted probability and statistical theory. Fischbein and Schnarch investigated the stability of students' misconceptions in probability across different age levels from grade 5 to college students. Surprisingly, they found three different outcomes: some misconceptions grew stronger with age, some grew weaker and one remained stable.

Vallecillos and Batanero presented a study on the persistence of conceptual errors in university students' understanding of levels of significance in tests of hypotheses. There are also parts on students' understanding and misunderstanding of concepts relating to probability and statistics in books on the teaching of statistics such as Green; Holmes; Hawkins, Jolliffe and Glickman. These attest to the importance of conceptual understanding as a goal of statistics instruction.

In summary, research on student learning in statistics shows that students have difficulties with statistical concepts and that they are often anxious about and have poor attitudes to learning statistics. The body of research on teaching and learning statistics is large, vibrant and growing. However, much of it still lacks the systematic methodology and theoretical foundations that are often of the most lasting value in any field of education. Further, a concern with describing teaching strategies or students' attitudes or their misconceptions, while indicating an awareness of the problems facing statistics education, does not offer a model for understanding the impasse–a first step towards alleviating it. My approach is to explore the issues underlying this impasse within a theoretical framework which relates students, their actions and the context.

ISSUES FOR EXPLORATION IN RESEARCH

The acknowledgment, in the education literature, of the prevalence of students' misinterpretations of key statistical concepts leads me to question how statistical thinking is embedded in academic settings. That is, how do students experience statistical reasoning in the academic setting and how does this reasoning relate to that of working statisticians in practical situations? According to one statistician the components of statistical thinking are: understanding the dynamics of the real world problem, moving towards a statistical model and using statistical tools.

If students in a university statistics course are not provided with experiences enabling them to understand the relationship between the statistical model and the "real life" situation, they may not be adequately equipped to use the statistical tools with which they are presented and so may be unable to appreciate the statistical reasoning process. For example a lack of understanding about variation could lead students either to trust implicitly their own intuition about data, without attempting to use statistical reasoning to evaluate the findings more critically, or to distrust universally any statistical statement–without the means to assess its reliability. It is important for me in this investigation to understand how students' concepts of statistical knowledge relate to the contexts in which they apply this knowledge. The ways that students experience learning statistics are also likely to contribute to the difficulties that many students have with the subject and to their objections to studying it. How students understand statistics is related to how they interpret it in the context in which it is presented.

This is an aspect of students' learning of statistics which has received scant attention in the research literature and, in my experience, is rarely accorded importance in the teaching of statistics courses in higher education. Skills and facts are easily identified in the design and assessment of statistics courses at university. However, students' conceptions of the subject matter as a whole, approaches to learning it and their perceptions and beliefs are rarely evaluated in any formal sense. These aspects of students' learning of Statistics are major topics for exploration in my project as it is a prime postulate of my research that these are critical dimensions of students' emerging statistical knowledge.

In learning Statistics, a student's thinking and problem solving is accompanied by affective elements: feelings, beliefs, desires and attitudes which will affect how long she will persist with a task or a problem, how easily distracted she is likely to be, how long she will remember facts and skills and other important factors which can facilitate or hinder her cognition. Affect is therefore of importance and interest in my investigation which seeks to understand learning Statistics from the activity theory perspective. In this framework, affective elements are not extraneous or secondary to intellectual processes but indivisible from them.

Vygotsky describes this inseparability of intellectual processes and affective elements as follows:

- Thought itself is engendered by motivation, *i.e.*, by our desires and needs, our interests and emotions. Behind every thought there is an affective-volitional tendency, which holds the answer to the last 'why' in the analysis of thinking. A true and full understanding of another's thought is possible only when we understand its affective-volitional basis.

Hence, the affective components of students' cognitive processes give those processes their "attitudinal colour". Statistics is included in the curricula of many disciplines at university in order that students may use it and appreciate its relevance to their chosen fields. In Western countries, statistical thinking is bound up with technology, commercial concerns and other matters of cultural importance. The social interests and preoccupations of a culture are not, however, automatically assimilated by individuals, but are monitored.

Students regulate their thinking and actions according to their evaluations of a learning task. What is culturally accepted may be personally abhorrent. There may even be a conflict between studying statistics and a student's perception of her own self realisation or values. For example, in the following survey response one student expressed the idea that real psychologists don't do mathematics.

She wrote:

- *I don't even see the point:* In psych why must maths infiltrate itself??? Studies have shown that those who have high maths abilities have low or poor communication and perception skills–shouldn't psychologists be exceptionally perceptive and able to communicate well? It seems that if there aren't silly numbers to justify things then they aren't plausible in our computer/maths/science promotive society.

Leont'ev makes an important distinction between meaning and personal sense, which relates to conflicts such as the one expressed by this student. To Leont'ev, meanings are connected with the reality of the outside objective world, the life of society. Leont'ev calls them the "crystallization of social experience". Personal sense, on the other hand, is connected with the reality of the person's own life and motives. That is, personal sense involves the incorporation of socially constituted meanings into the psychology of the individual.

Personal sense, according to Leont'ev "does not have its own 'supraindividual', 'nonpsychological' existence". Sometimes, there is a mismatch between societal and personal meaning. According to Leont'ev, it is affect that signals to the individual the fit between the two.

Lerman argues that the valuing of decontextualised, intellectual processes, divorced from personal and social elements, is expressive of oppressive discourse. It is this privileging of abstract thought, such as academic mathematics, that is disempowering for some students. Sierpinska and Lerman refer to the vested interest mathematicians have in maintaining the status of mathematics in society. This idea was passionately expressed by a participant in my pilot study for Study Two.

She wrote:

- Maths is an exercise in agony, because the people who teach it make one feel as though maths belongs in a higher plane of evolution. Even though the number system is for everyone, and the concepts are there for everyone, the feeling that you do not deserve to know anything runs rampant. Maths, in short, is a lofty pain and a real headache to study.

In summary, the issues that I am concerned with in this project relate to the ways that students orient themselves to and interpret the task of learning Statistics and to how personal dimensions and individual actions are linked to the wider social, institutional and cultural arenas surrounding them.

RESEARCH DESIGN AND APPROACH

The activity theory perspective suggests that a focus on teaching, unrelated to students' perceptions about their own learning and its context, is unlikely to be successful in changing outcomes. My approach focuses on the experience of learning Statistics at university from the point of view of the student. Consistent with my theoretical framework based on the ideas of Leont'ev and Vygotsky, I explore the network of relationships between learners, the subject matter and context.

My investigation took place at a large and traditional metropolitan university, the oldest in Sydney. The participants were second year Psychology students, all of whom were required to study Statistics. Statistics is introduced in first year Psychology at The University of Sydney, but it is as a compulsory component of the second year course that it plays a major role, contributing one quarter of the final assessment mark. The instruments I used included participant observation, interviews and surveys. My project consists of two investigations.

Study One is entitled "Exploring Mature Students' Learning of Statistics". The context for this study was the Mathematics Learning Centre where I work. The participants in this study were of great interest to me as an educator concerned with ameliorating students' difficulties with Statistics. Study Two is called "Understanding University Students Learning Statistics as a Service Course to Psychology".

In this study, I widened the context of my inquiry to include the general Statistics class. For my investigations, I drew on a triad of research tools– naturalistic inquiry, procedures based on the phenomenographic method and quantitative methods for describing data and exploring patterns and relationships. Among the instruments I developed was the Approaches to Learning Statistics Questionnaire.

PSYCHOLOGY STUDENTS AT THE MATHEMATICS LEARNING CENTRE

In Study One I utilise a naturalistic perspective to draw on the rich accounts of five students who were studying Psychology after many years away from the educational setting. My aim is to tell the story of these students in a way that is generalisable to others–that is, I try to uncover recognisable insights. I was in no way involved with the formal assessment of these students and worked very intensively and closely with them. Hence I was able to develop a close relationship with the participants of Study One and had more access to their perceptions and feelings than is usually the case for university educators in statistics.

My analysis is based mainly on interview data, complemented by written expressions from various sources. The five participants were selected "purposively" for their range of experiences and attributes. As is often the case with qualitative research, the data for this study consisted of descriptions, direct quotations and excerpts, obtained by close psychological contact with those being studied.

The study is an example of action research in that my actions as a researcher were inseparable from those designed to assist the students in their learning of Statistics. As I have outlined in Gordon this means that the methodology for the study developed through the "action research spiral".

They describe this spiral as recurring cycles of:

- Planning, acting, observing and reflecting, with each of these activities being systematically and self-critically implemented and interrelated.

VIEW OF PSYCHOLOGY STUDENTS LEARNING STATISTICS

Study Two was grounded in Study One and broadened the perspective on students learning Statistics. The participants of Study Two were Psychology II Students who took part in a survey during their Statistics lecture, near the end of the first semester. The survey was completed during 20 minutes of the lecture time and was the major source of data for this study. Other data for this study included assessment results and vignettes from interviews with students selected from those who agreed to contribute further to my research.

Background of Phenome Nography as a Research Tool

One major focus of Study Two was to elicit students' conceptions of the subject matter, in the context in which they were studying it. To do this I adopted a phenomenographic approach. The term phenomenography was first used by Ference Marton to describe programmes of research which had the common aim of describing peoples' conceptions.

The results of phenomenographic analysis are categories of description of the qualitatively different ways students conceive of the phenomenon being investigated. This focus on describing conceptions is not the sole aim of the investigations in which phenomenography is used as a research tool. Rather it is learning and teaching that are being investigated based on the focusing of conceptions. Conceptions are the lens through which the phenomenographic researcher views learning.

Marton describes phenomenography as a method:

- For investigating the qualitatively different ways in which people experience, conceptualise, perceive and understand various aspects of, and phenomena in, the world around them.

Phenomenography was developed within the framework of education research at the University of Gothenburg, Sweden. It is concerned with the relations between people and the objects of their perceptions or content of their thoughts. Marton calls phenomenography a "research specialization", that is, a combination of a research orientation and a research approach to describing and comparing conceptions and understandings.

The research orientation emphasises the importance of description and the use of categories as forms of expressing the conceptions. The research approach is characterised by the open explorative form of data collection and the interpretative analysis of data. Säljö, a forerunner in the development of the phenomenographic approach, comments that in more recent studies, the term "way of experiencing" is used in preference to the term "conception".

This emphasises that:

- The prime interest of phenomenographic research is in finding and delimiting the variation in ways of experiencing reality.

Hence, in deriving categories of description for students' conceptions of Statistics, I am attempting to find relationships between the students and the subject matter which express their ways of experiencing the subject, Statistics, as they were learning it. Svensson argues that underpinning the phenomenographic specialisation is an assumption that knowledge is fundamentally a "question of meaning in a social and cultural context". Such an approach views phenomena systemically and avoids separating person and context.

As Säljö in an editorial of Learning and Instruction, concludes:

- Human experiences are inescapably cultural in nature, and learning and growth take place within cultural boundaries.

This view of the intertwining of culture, mind and action is consistent with a Vygotskian assumption that there is no duality between self and context; between thinking and acting. The phenomenographic view of knowledge is that it is relational, created through human thinking and activity which relates to the external world–a relation between thought and social and cultural life. Hence my choice of phenomenography as one of my research tools is consistent with the theoretical perspective with which I am framing this investigation.

The relational view of learning furnished by phenomenography has formed the basis of much of the research into student learning in higher education. Phenomenography has been fruitful in exploring students' conceptions and understanding in a number of topics, including physics; literature interpretation; essay writing and computer programming. In summary, this research indicates that learners' experiences should be considered as involving the ways students relate to the learning environment, their goals and intentions, as well as their learning strategies. My colleagues and I have found phenomenographic methods useful to explore the qualitatively different ways in which students experience learning mathematics. Two studies were conducted.

In the first study we investigated the conceptions of mathematics and approaches to learning it of students entering university. In the second investigation we explored these students' conceptions of a fundamental mathematical concept, namely that of a function. Traditionally, phenomenographic studies are built around interviews. However, in these investigations, my colleagues and I extended previous research by constructing categories from the written responses of a large number of students. This developed phenomenography as a research tool. In Study Two, I built on our previous phenomenographic research by looking at the variation in the way second year Psychology students conceived of the nature of the subject matter, Statistics. The phenomenographic perspective provided my basis for identifying categories of description of students' conceptions. These revealed qualitatively different ways of experiencing Statistics. In much research, the categories of description or variables are imposed on the data by an expert researcher.

In Study Two, in keeping with phenomenographic practice, the categories of description were constituted in relation to the data–from interpretations of the survey responses. This does not imply that I, as researcher, viewed the data free of ideas based on prior experience or theoretical orientation.

Indeed:

- The act of categorisation is by its very nature subjective.

Further, my research is deliberately set in the framework of activity theory. However, instead of characterising students' responses to the three open-ended questions in the survey by means of an established taxonomy, the categories emerged in the context of the study. By keeping the questions as general as possible, I tried not to shape students' responses. Rather I tried to explore the students' points of view. This enhances the ethnographic validity of this research, that is, the sense of the research in terms of the lives and awareness of the participants in it.

The phenomenographic method analyses meanings constituted within a particular context and content domain. My categories were derived in a particular setting. However, they and their structure as an organised set, may prove useful in contexts other than the one from which they emerged.

In accord with the phenomenographic approach:

- The categories of description are supposed to be replicable, but theway in which they were found is not.

Conceptions may be expressed in different forms of action but they are most accessible through language. Variation in conceptions, expressed by people, immediately brings in the cultural and social context in which these conceptions are reported. The students' reports on their ways of conceiving Statistics, although limited to the students' awareness and abilities in written verbal expression, provided me with a rich source of information for characterising these experiences. This experiential perspective is referred to as a "second order" perspective by Marton.

He contrasts this with phenomenology, based on Husserl's imperative to return to the essence of a phenomenon, through immediate experience and free from conceptual thought–a first order enterprise. Phenomenography is a perspective on how things appear to people–how phenomena are experienced, rather than what they are "in reality" or whether or not such a reality can be described.

Phenomenography was derived from empirical and pragmatic research in education. Its commitment to philosophical foundations is therefore not always clear. The procedures defining its use are neither prescriptive nor indisputable. Indeed, Hasselgren and Beach argue that "there is no genuine consensus method of phenomenography".

To them:

- Phenomenography is research which is simply concerned with how things are understood, the experiences of the process of formation of understandings at individual levels, and their distributions in specific collectivities.

While my way of applying this research tool is individual, I have tried to maintain the essence of phenomenographic research as a way of understanding the multiple realities of students. That is, my appropriation of Marton's term signifies this approach. I view phenomenography as a research tool which bridges the gap between empirical, analytic methods, which are recognised by the research community, and students' subjective awareness of their own diverse conceptions and experiences.

Introduction to Analysis of Students' Approaches to Learning Statistics

I analysed students' approaches to learning Statistics by means of a questionnaire, the Approaches to Learning Statistics Questionnaire. My questionnaire was modified from the Approaches to learning Mathematics Questionnaire which, in turn, was derived by my colleagues and myself from the Study Process Questionnaire. Biggs refers to three approaches to learning: deep, surface and achieving. He differentiated each of these into two aspects– the motive, or intention, and the strategy. Hence, in the Study Process Questionnaire the items are divided into those referring to motive or intention and those referring to strategy.

Each motive-strategy combination defines a distinct approach to learning– deep, surface or achieving. Early work by Biggs, on samples of students from universities and Colleges of Advanced Education, indicated that the achieving approach was equally related to the surface and deep approaches in tertiary students.

The surface and deep approaches, however, are referred to by Biggs as:

- Independent ways in which students may become involved in learning.

In this project I was concerned with deep and surface approaches to learning. Achieving approaches were of less theoretical interest to me and were not investigated. Considerable research has been done on the Study Process Questionnaire. For example, a literature review on approaches to studying indicates that mature students seem more likely than younger students to adopt deep approaches to learning, while surface approaches decline with age. However, completion rates and possible bias in these studies, for example, in the postal returns of surveys in Biggs' investigations, renders this finding problematic.

Richardson also reviewed the literature on cultural specificity of approaches to studying in higher education. Investigations using the Study Process Questionnaire in countries as diverse as Australia, Nepal, Britain and Nigeria yielded a similar two factor structure for this questionnaire–one indicating a deep approach and the other a surface approach. The research also suggests that orientations towards understanding the meaning of learning materials, that is, deep approaches to learning, reflect a consistent agreement across different cultures with regard to the goals of higher education.

On the other hand, this research indicates that patterns of responses on the subscales measuring surface approaches to learning, are more distinctive to differing cultural contexts. Concern has been expressed about the factor structure underlying Biggs' differentiation of the Study Process Questionnaire into motive and strategy components. In my analysis of the Approaches to Learning Statistics Questionnaire I did not find evidence for a breakdown into intention and strategy factors. I therefore analysed the items of the ALSQ in terms of two subscales: one measuring a general surface approach and the other a deep approach to learning Statistics. Biggs characterises a surface approach and a deep approach to an academic task by the following.

A surface approach is distinguished by:

- A view of the task as imposed, a demand to be met;
- A fragmented view of the task–parts are unrelated to each other and the task is unrelated to other tasks;
- Worries about the time taken by the task;
- A lack of recognition of any personal meaning in the task;
- Reliance on memorisation and ways of reproducing the details.

A deep approach reflects:

- Interest and enjoyment in the task;
- A search for underlying meaning;
- Efforts to make the learning personally meaningful–relating task to personal experience or to the real world;
- Holistic approach–relating parts of the task to each other and to other knowledge;
- Theorising about the task.

The surface approach and the deep approach, as defined by Biggs, correspond closely with the surface and deep levels of processing distinguished by Marton and Säljö. Later, Marton and Säljö referred to these as surface and deep approaches to learning. In contrast to Biggs, however, Marton and Säljö see surface and deep approaches as the opposite ends of a single continuum, rather than as independent ways of learning. Marton and Säljö showed that students would adopt one of two methods of processing information according to their intentions.

If their aim was merely to memorise and reproduce material, or, as Biggs puts it:

- To display the symptoms of having learned,

They would adopt a surface level of processing; if they wanted to maximise their understanding of the underlying meaning, they would adopt a deep strategy. Hence Marton and Säljö inferred a relationship between strategy and intention. What the student intends to gain from the learning determines the strategy used. In recent research, Marton, Watkins and Tang make the point that while a surface approach to learning is often associated with rote learning, this is not what characterises it.

- Rather, a surface approach is characterised by a focus on the learning material or task in itself and not, as would be the case for the deep approach, on the meaning or purpose underlying it.

Their ideas accord with Leont'ev's proposition that what distinguishes actions are the goals driving them.

INVESTIGATING INDIVIDUALS IN SOCIALLY SIGNIFICANT PRACTICES

My project is an example of what Chaiklin describes as "investigating individuals in socially significant practices".

These are studies in a "yet-to -be-embodied" research tradition that:

- Try and develop an account of the actions of individuals participating in a societally significant practice while it is occurring, by an analysis that locates the practice in a social, societal and/or historical perspective.

Chaiklin proposes that many different theoretical perspectives and methods could contribute to such research. He summarises such projects as having five common characteristics, which I assert are fulfilled by my project.

- Firstly, these studies take concrete, meaningful, societal practices as the direct object of study. That is, the setting for the study is an actual example of human practice and would be so whether or not it was the object of study. In my research I have tried to get as close as possible to the participants and interpret situations which were as typical as possible for students learning Statistics.
- Secondly, the practice takes place in a recognisable societal institution, in this case an institution of higher learning.
- Thirdly, such studies have a definite theoretical interest. Insights into the practice of learning Statistics are important elements of my research but so are insights into and extensions of the activity theory framework in which I have interpreted this practice.
- Fourthly, there is the idea that knowledge and actions are social as well as individual. I draw on Vygotsky's and Leont'ev's idea of "co-knowing" to illustrate this notion–statistical knowledge is socially created and re-created through the consensus of practitioners. This collective development of knowledge is facilitated through tools, for example, by the easy and quick exchange of ideas through electronic media. In my university, Statistics is continually being transformed by each lecturer and committee involved in organising the curriculum and setting the examinations. Students come to know Statistics through interacting with peers and teachers, as well as through their own reflections and actions. Socially organised constraints and resources surround individual actions.
- *Finally, the particular practice selected for such studies has:*
  - Significant consequences for the people participating in these practices.

Learning Statistics can change the lives of those participating in this practice. In some cases, the outcomes could be instrumental in determining students' academic and career paths. For example, students' examination results in Statistics could determine whether they would attain a degree or be accepted into Psychology Honours. Learning Statistics can also powerfully affect students' personal development as was the case for some of the participants of Study One.

In this project I have tried to piece together practical and theoretical perspectives to contribute to the building of a research framework–to an understanding of the relations between theory and practice in education.

2

Statistical Regression Analysis

LINEAR DISCRIMINANT ANALYSIS

Linear discriminant analysis (LDA) and the related Fisher's linear discriminant are methods used in statistics, pattern recognition and machine learning to find a linear combination of features which characterize or separate two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

LDA is closely related to ANOVA (analysis of variance) and regression analysis, which also attempt to express one dependent variable as a linear combination of other features or measurements. In the other two methods however, the dependent variable is a numerical quantity, while for LDA it is a categorical variable (*i.e.*, the class label). Logistic regression and probit regression are more similar to LDA, as they also explain a categorical variable. These other methods are preferable in applications where it is not reasonable to assume that the independent variables are normally distributed, which is a fundamental assumption of the LDA method.

LDA is also closely related to principal component analysis (PCA) and factor analysis in that both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account any difference in class, and factor analysis builds the feature combinations based on differences rather than similarities. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique: a distinction between independent variables and dependent variables (also called criterion variables) must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities. When dealing with categorical independent variables, the equivalent technique is discriminant correspondence analysis.

PRACTICAL USE

In practice, the class means and covariances are not known. They can, however, be estimated from the training set. Either the maximum likelihood estimate or the maximum a posteriori estimate may be used in place of the exact value in the above equations. Although the estimates of the covariance may be considered optimal in some sense, this does not mean that the resulting discriminant obtained by substituting these values is optimal in any sense, even if the assumption of normally distributed classes is correct.

Another complication in applying LDA and Fisher's discriminant to real data occurs when the number of observations of each sample does not exceed the number of samples. In this case, the covariance estimates do not have full rank, and so cannot be inverted. There are a number of ways to deal with this. One is to use a pseudo inverse instead of the usual matrix inverse in the above formulae. However, better numeric stability may be achieved by first projecting the problem onto the subspace spanned by $Ó_b$. Another strategy to deal with small sample size is to use a Shrinkage estimator of the covariance matrix, which can be expressed mathematically as

$$C_{new} = (1 - \lambda) C + \lambda I$$

where *I* is the identity matrix, and λ is the *shrinkage intensity* or *regularisation parameter*. This leads to the framework of regularized discriminant analysis or shrinkage discriminant analysis. Also, in many practical cases linear discriminants are not suitable. LDA and Fisher's discriminant can be extended for use in non-linear classification via the kernel trick. Here, the original observations are effectively mapped into a higher dimensional non-linear space. Linear classification in this non-linear space is then equivalent to non-linear classification in the original space. The most commonly used example of this is the kernel Fisher discriminant. LDA can be generalized to multiple discriminant analysis, where *c* becomes a categorical variable with *N* possible states, instead of only two. Analogously, if the class-conditional densities are normal with shared covariances, the sufficient statistic for are the values of *N* projections, which are the subspace spanned by the *N* means, affine projected by the inverse covariance matrix.

These projections can be found by solving a generalized eigenvalue problem, where the numerator is the covariance matrix formed by treating the means as the samples, and the denominator is the shared covariance matrix.

APPLICATIONS

In addition to the examples given below, LDA is applied in positioning and product management.

BANKRUPTCY PREDICTION

In bankruptcy prediction based on accounting ratios and other financial variables, linear discriminant analysis was the first statistical method applied to systematically explain which firms entered bankruptcy vs. survived. Despite limitations including known nonconformance of accounting ratios to the normal distribution assumptions of LDA, Edward Altman's 1968 model is still a leading model in practical applications.

FACE RECOGNITION

In computerised face recognition, each face is represented by a large number of pixel values. Linear discriminant analysis is primarily used here to reduce the number of features to a more manageable number before classification. Each of the new dimensions is a linear combination of pixel values, which form a template. The linear combinations obtained using Fisher's linear discriminant are called *Fisher faces*, while those obtained using the related principal component analysis are called *eigenfaces*.

MARKETING

In marketing, discriminant analysis was once often used to determine the factors which distinguish different types of customers and/or products on the basis of surveys or other forms of collected data. Logistic regression or other methods are now more commonly used.

The use of discriminant analysis in marketing can be described by the following steps:

1. Formulate the problem and gather data — Identify the salient attributes consumers use to evaluate products in this category — Use quantitative marketing research techniques (such as surveys) to collect data from a sample of potential customers concerning their ratings of all the product attributes. The data collection stage is usually done by marketing research professionals. Survey questions ask the respondent to rate a product from one to five (or 1 to 7, or 1 to 10) on a range of attributes chosen by the researcher. Anywhere from five to twenty attributes are chosen. They could include things like: ease of use, weight, accuracy, durability, colourfulness, price, or size. The attributes chosen will vary depending on the product being studied. The same question is asked about all the products in the study. The data for multiple products is codified and input into a statistical programme such as R, SPSS or SAS. (This step is the same as in Factor analysis).
2. Estimate the Discriminant Function Coefficients and determine the statistical significance and validity — Choose the appropriate discriminant analysis method. The direct method involves estimating the discriminant function so that all the predictors are assessed simultaneously. The stepwise method enters the predictors sequentially.

The two-group method should be used when the dependent variable has two categories or states. The multiple discriminant method is used when the dependent variable has three or more categorical states. Use Wilks's Lambda to test for significance in SPSS or F stat in SAS. The most common method used to test validity is to split the sample into an estimation or analysis sample, and a validation or holdout sample. The estimation sample is used in constructing the discriminant function. The validation sample is used to construct a classification matrix which contains the number of correctly classified and incorrectly classified cases. The percentage of correctly classified cases is called the hit ratio.

3. Plot the results on a two dimensional map, define the dimensions, and interpret the results. The statistical programme (or a related module) will map the results. The map will plot each product (usually in two dimensional space). The distance of products to each other indicate either how different they are. The dimensions must be labelled by the researcher. This requires subjective judgement and is often very challenging.

LINEAR REGRESSION

In statistics, linear regression is an approach to modelling the relationship between a scalar dependent variable *y* and one or more explanatory variables denoted *X*. The case of one explanatory variable is called *simple regression*. More than one explanatory variable is *multiple regression*. (This in turn should be distinguished from *multivariate linear regression*, where multiple correlated dependent variables are predicted, rather than a single scalar variable.)

In linear regression, data are modelled using linear predictor functions, and unknown model parameters are estimated from the data. Such models are called *linear models*. Most commonly, linear regression refers to a model in which the conditional mean of *y* given the value of *X* is an affine function of *X*. Less commonly, linear regression could refer to a model in which the median, or some other quantile of the conditional distribution of *y* given *X* is expressed as a linear function of *X*. Like all forms of regression analysis, *linear regression* focuses on the conditional probability distribution of *y* given *X*, rather than on the joint probability distribution of *y* and *X*, which is the domain of multivariate analysis.

Linear regression was the first type of regression analysis to be studied rigorously, and to be used extensively in practical applications. This is because models which depend linearly on their unknown parameters are easier to fit than models which are non-linearly related to their parameters and because the statistical properties of the resulting estimators are easier to determine.

Linear regression has many practical uses. Most applications of linear regression fall into one of the following two broad categories:

- If the goal is prediction, or forecasting, linear regression can be used to fit a predictive model to an observed data set of *y* and *X* values.

After developing such a model, if an additional value of X is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

- Given a variable y and a number of variables $X_1,..., X_p$ that may be related to y, linear regression analysis can be applied to quantify the strength of the relationship between $y$and the X_j, to assess which X_j may have no relationship with y at all, and to identify which subsets of the X_j contain redundant information about y.

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as withleast absolute deviations regression), or by minimizing a penalized version of the least squaresloss function as in ridge regression. Conversely, the least squares approach can be used to fit models that are not linear models. Thus, while the terms "least squares" and "linear model" are closely linked, they are not synonymous.

Some remarks on terminology and general use:

- Is called the *regressand*, *exogenous variable*, *response variable*, *measured variable*, or *dependent variable*. The decision as to which variable in a data set is modeled as the dependent variable and which are modeled as the independent variables may be based on a presumption that the value of one of the variables is caused by, or directly influenced by the other variables. Alternatively, there may be an operational reason to model one of the variables in terms of the others, in which case there need be no presumption of causality.
- Are called *regressors*, *endogenous variables*, *explanatory variables*, *covariates*, *input variables*, *predictor variables*, or *independent variables* The matrix is sometimes called the design matrix.
- Usually a constant is included as one of the regressors. For example we can take $x_{i1} = 1$ for $i = 1,..., n$. The corresponding element of β is called the *intercept*. Many statistical inference procedures for linear models require an intercept to be present, so it is often included even if theoretical considerations suggest that its value should be zero.
- Sometimes one of the regressors can be a non-linear function of another regressor or of the data, as in polynomial regression and segmented regression. The model remains linear as long as it is linear in the parameter vector β.
- The regressors x_{ij} may be viewed either as random variables, which we simply observe, or they can be considered as predetermined fixed values which we can choose. Both interpretations may be appropriate in different cases, and they generally lead to the same estimation procedures; however different approaches to asymptotic analysis are used in these two situations.
- Is a p-dimensional *parameter vector*. Its elements are also called *effects*, or *regression coefficients*. Statistical estimation andinference in linear regression focuses on β.

- Is called the *error term*, *disturbance term*, or *noise*. This variable captures all other factors which influence the dependent variable y_i other than the regressors x_i. The relationship between the error term and the regressors, for example whether they are correlated, is a crucial step in formulating a linear regression model, as it will determine the method to use for estimation.

ASSUMPTIONS

Standard linear regression models with standard estimation techniques make a number of assumptions about the predictor variables, the response variables and their relationship. Numerous extensions have been developed that allow each of these assumptions to be relaxed (*i.e.*, reduced to a weaker form), and in some cases eliminated entirely. Some methods are general enough that they can relax multiple assumptions at once, and in other cases this can be achieved by combining different extensions. Generally these extensions make the estimation procedure more complex and time-consuming, and may also require more data in order to get an accurate model.

The following are the major assumptions made by standard linear regression models with standard estimation techniques (e.g., ordinary least squares):

- *Weak exogeneity:* This essentially means that the predictor variables x can be treated as fixed values, rather than random variables. This means, for example, that the predictor variables are assumed to be error-free, that is they are not contaminated with measurement errors. Although not realistic in many settings, dropping this assumption leads to significantly more difficult errors-in-variables models.
- *Linearity:* This means that the mean of the response variable is a linear combination of the parameters (regression coefficients) and the predictor variables. Note that this assumption is much less restrictive than it may at first seem. Because the predictor variables are treated as fixed values, linearity is really only a restriction on the parameters. The predictor variables themselves can be arbitrarily transformed, and in fact multiple copies of the same underlying predictor variable can be added, each one transformed differently. This trick is used, for example, in polynomial regression, which uses linear regression to fit the response variable as an arbitrary polynomial function (up to a given rank) of a predictor variable. This makes linear regression an extremely powerful inference method. In fact, models such as polynomial regression are often "too powerful", in that they tend tooverfit the data. As a result, some kind of regularization must typically be used to prevent unreasonable solutions coming out of the estimation process. Common examples are ridge regression and lasso regression. Bayesian linear regression can also be used, which by its nature is more or less immune to the problem of overfitting. (In fact, ridge regression and lasso regression can both be viewed as special cases of Bayesian linear regression, with particular types of prior distributions placed on the regression coefficients.)

- *Constant variance:* This means that different response variables have the same variance in their errors, regardless of the values of the predictor variables. In practice this assumption is invalid (*i.e.*, the errors are heteroscedastic) if the response variables can vary over a wide scale. In order to determine for heterogeneous error variance, or when a pattern of residuals violates model assumptions of homoscedasticity (error is equally variable around the 'best-fitting line' for all points of x), it is prudent to look for a "fanning effect" between residual error and predicted values. This is to say there will be a systematic change in the absolute or squared residuals when plotted against the predicting outcome. Error will not be evenly distributed across the regression line. Heteroscedasticity will result in the averaging over of distinguishable variances around the points to get a single variance that is inaccurately representing all the variances of the line. In effect, residuals appear clustered and spread apart on their predicted plots for larger and smaller values for points along the linear regression line, and the mean squared error for the model will be wrong. Typically, for example, a response variable whose mean is large will have a greater variance than one whose mean is small. For example, a given person whose income is predicted to be $100,000 may easily have an actual income of $80,000 or $120,000 (a standard deviation of around $20,000), while another person with a predicted income of $10,000 is unlikely to have the same $20,000 standard deviation, which would imply their actual income would vary anywhere between -$10,000 and $30,000. (In fact, as this shows, in many cases – often the same cases where the assumption of normally distributed errors fails – the variance or standard deviation should be predicted to be proportional to the mean, rather than constant.) Simple linear regression estimation methods give less precise parameter estimates and misleading inferential quantities such as standard errors when substantial heteroscedasticity is present. However, various estimation techniques (*e.g.*, weighted least squares and heteroscedasticity-consistent standard errors) can handle heteroscedasticity in a quite general way. Bayesian linear regression techniques can also be used when the variance is assumed to be a function of the mean. It is also possible in some cases to fix the problem by applying a transformation to the response variable (*e.g.*, fit the logarithm of the response variable using a linear regression model, which implies –that the response variable has a log-normal distribution rather than a normal distribution).
- *Independence of errors:* This assumes that the errors of the response variables are uncorrelated with each other. (Actual statistical independence is a stronger condition than mere lack of correlation and is often not needed, although it can be exploited if it is known to hold.) Some methods (*e.g.*, generalized least squares) are capable of handling correlated errors, although they typically require significantly more data

unless some sort of regularization is used to bias the model towards assuming uncorrelated errors. Bayesian linear regression is a general way of handling this issue.

- *Lack of multicollinearity in the predictors:* For standard least squares estimation methods, the design matrix *X* must have fullcolumn rank *p*, *i.e.*, be invertible; otherwise, we have a condition known as multicollinearity in the predictor variables. This can be triggered by having two or more perfectly correlated predictor variables (*e.g.*, if the same predictor variable is mistakenly given twice, either without transforming one of the copies or by transforming one of the copies linearly). It can also happen if there is too little data available compared to the number of parameters to be estimated (*e.g.*, fewer data points than regression coefficients). In the case of multicollinearity, the parameter vector β will be non-identifiable — it has no unique solution. At most we will be able to identify some of the parameters, *i.e.*, narrow down its value to some linear subspace of R^p. Methods for fitting linear models with multicollinearity have been developed; some require additional assumptions such as "effect sparsity" — that a large fraction of the effects are exactly zero. Note that the more computationally expensive iterated algorithms for parameter estimation, such as those used in generalized linear models, do not suffer from this problem — and in fact it's quite normal to when handling categorically-valued predictors to introduce a separate indicator variable predictor for each possible category, which inevitably introduces multicollinearity.

Beyond these assumptions, several other statistical properties of the data strongly influence the performance of different estimation methods:

- The statistical relationship between the error terms and the regressors plays an important role in determining whether an estimation procedure has desirable sampling properties such as being unbiased and consistent.
- The arrangement, or probability distribution of the predictor variables *x* has a major influence on the precision of estimates of β.Sampling and design of experiments are highly-developed subfields of statistics that provide guidance for collecting data in such a way to achieve a precise estimate of β.

INTERPRETATION

A fitted linear regression model can be used to identify the relationship between a single predictor variable $x_j$and the response variable *y* when all the other predictor variables in the model are "held fixed". Specifically, the interpretation of $\hat{a}_j$ is the expected change in *y* for a one-unit change in x_j when the other covariates are held fixed—that is, the expected value of the partial derivativeof *y* with respect to x_j. This is sometimes called the *unique effect* of x_j on *y*. In contrast, the *marginal effect* of x_j on *y* can be assessed using a correlation coefficient or simple linear regression model relating $x_j$to *y*; this effect is the total derivative of y with respect tox_j. Care must be taken when interpreting regression results, as some of the regressors may not allow for marginal changes (such as dummy variables, or the intercept term), while others cannot be held fixed (recall the example from the introduction: it would be impossible to "hold t_i fixed" and at the same time change the value of t_i^2). It is possible that the unique effect can be nearly zero even when the marginal effect is large.

This may imply that some other covariate captures all the information in x_j, so that once that variable is in the model, there is no contribution of x_j to the variation in y. Conversely, the unique effect of x_j can be large while its marginal effect is nearly zero. This would happen if the other covariates explained a great deal of the variation of y, but they mainly explain variation in a way that is complementary to what is captured by x_j. In this case, including the other variables in the model reduces the part of the variability of y that is unrelated to x_j, thereby strengthening the apparent relationship with x_j.

The meaning of the expression "held fixed" may depend on how the values of the predictor variables arise. If the experimenter directly sets the values of the predictor variables according to a study design, the comparisons of interest may literally correspond to comparisons among units whose predictor variables have been "held fixed" by the experimenter. Alternatively, the expression "held fixed" can refer to a selection that takes place in the context of data analysis. In this case, we "hold a variable fixed" by restricting our attention to the subsets of the data that happen to have a common value for the given predictor variable. This is the only interpretation of "held fixed" that can be used in an observational study.

The notion of a "unique effect" is appealing when studying a complex system where multiple interrelated components influence the response variable. In some cases, it can literally be interpreted as the causal effect of an intervention that is linked to the value of a predictor variable. However, it has been argued that in many cases multiple regression analysis fails to clarify the relationships between the predictor variables and the response variable when the predictors are correlated with each other and are not assigned following a study design.

EXTENSIONS

Numerous extensions of linear regression have been developed, which allow some or all of the assumptions underlying the basic model to be relaxed.

SIMPLE AND MULTIPLE REGRESSION

The very simplest case of a single scalar predictor variable x and a single scalar response variable y is known as *simple linear regression*. The extension to multiple and/or vector-valued predictor variables (denoted with a capital X) is known as *multiple linear regression*. Nearly all real-world regression models involve multiple predictors, and basic descriptions of linear regression are often phrased in terms of the multiple regression model. Note, however, that in these cases the response variable y is still a scalar.

GENERAL LINEAR MODELS

The general linear model considers the situation when the response variable Y is not a scalar but a vector. Conditional linearity of $E(y|x) = Bx$ is still assumed, with a matrix B replacing the vector β of the classical linear regression model. Multivariate analogues of OLS and GLS have been developed.

HETEROSKEDASTIC MODELS

Various models have been created that allow for heteroskedasticity, *i.e.*, the errors for different response variables may have different variances. For example, weighted least squares is a method for estimating linear regression models when the response variables may have different error variances, possibly with correlated errors. #Weighted linear least squares, and generalized least squares.) Heteroscedasticity-consistent standard errors is an improved method for use with uncorrelated but potentially heteroskedastic errors.

GENERALIZED LINEAR MODELS

Generalized linear models (GLM's) are a framework for modeling a response variable y that is bounded or discrete. This is used, for example:

- When modeling positive quantities (*e.g.*, prices or populations) that vary over a large scale — which are better described using askewed distribution such as the log-normal distribution or Poisson distribution (although GLM's are not used for log-normal data, instead the response variable is simply transformed using the logarithm function);
- When modeling categorical data, such as the choice of a given candidate in an election (which is better described using a Bernoulli distribution/ binomial distribution for binary choices, or a categorical distribution/ multinomial distribution for multi-way choices), where there are a fixed number of choices that cannot be meaningfully ordered;
- When modeling ordinal data, *e.g.*, ratings on a scale from 0 to 5, where the different outcomes can be ordered but where the quantity itself may not have any absolute meaning (*e.g.*, a rating of 4 may not be "twice as good" in any objective sense as a rating of 2, but simply indicates that it is better than 2 or 3 but not as good as 5).

Generalized linear models allow for an arbitrary *link function g* that relates the mean of the response variable to the predictors, *i.e.*, $y = g(\hat{a}2\ x) + \varepsilon$. The link function is often related to the distribution of the response, and in particular it typically has the effect of transforming between the range of the linear predictor and the range of the response variable.

Some common examples of GLM's are:

- Poisson regression for count data.
- Logistic regression and probit regression for binary data.
- Multinomial logistic regression and multinomial probit regression for categorical data.
- Ordered probit regression for ordinal data.

Single index models allow some degree of nonlinearity in the relationship between *x* and *y*, while preserving the central role of the linear predictor *â*2 *x* as in the classical linear regression model. Under certain conditions, simply applying OLS to data from a single-index model will consistently estimate β up to a proportionality constant.

HIERARCHICAL LINEAR MODELS

Hierarchical linear models (or *multilevel regression*) organizes the data into a hierarchy of regressions, for example where *A* is regressed on *B*, and *B* is regressed on *C*. It is often used where the data have a natural hierarchical structure such as in educational statistics, where students are nested in classrooms, classrooms are nested in schools, and schools are nested in some administrative grouping such as a school district. The response variable might be a measure of student achievement such as a test score, and different covariates would be collected at the classroom, school, and school district levels.

ERRORS-IN-VARIABLES

Errors-in-variables models (or "measurement error models") extend the traditional linear regression model to allow the predictor variables *X* to be observed with error. This error causes standard estimators of β to become biased. Generally, the form of bias is an attenuation, meaning that the effects are biased towards zero.

OTHERS

- In Dempster–Shafer theory, or a linear belief function in particular, a linear regression model may be represented as a partially swept matrix, which can be combined with similar matrices representing observations and other assumed normal distributions and state equations. The combination of swept or unswept matrices provides an alternative method for estimating linear regression models.

ESTIMATION METHODS

A large number of procedures have been developed for parameter estimation and inference in linear regression. These methods differ in computational simplicity of algorithms, presence of a closed-form solution, robustness with respect to heavy-tailed distributions, and theoretical assumptions needed to validate desirable statistical properties such as consistency and asymptotic efficiency. Some of the more common estimation techniques for linear regression are summarized below.

MAXIMUM-LIKELIHOOD ESTIMATION AND RELATED TECHNIQUES

- Maximum likelihood estimation can be performed when the distribution of the error terms is known to belong to a certain parametric family $f_{è}$

of probability distributions. When $f_{è}$ is a normal distribution with mean zero and variance è, the resulting estimate is identical to the OLS estimate. GLS estimates are maximum likelihood estimates when ε follows a multivariate normal distribution with a known covariance matrix.

- Ridge regression, and other forms of penalized estimation such as Lasso regression, deliberately introduce bias into the estimation of β in order to reduce the variability of the estimate. The resulting estimators generally have lower mean squared error than the OLS estimates, particularly when multicollinearity is present. They are generally used when the goal is to predict the value of the response variable *y* for values of the predictors *x* that have not yet been observed. These methods are not as commonly used when the goal is inference, since it is difficult to account for the bias.
- Least absolute deviation (LAD) regression is a robust estimation technique in that it is less sensitive to the presence of outliers than OLS (but is less efficient than OLS when no outliers are present). It is equivalent to maximum likelihood estimation under aLaplace distribution model for ε.
- Adaptive estimation. If we assume that error terms are independent from the regressors, the optimal estimator is the 2-step MLE, where the first step is used to non-parametrically estimate the distribution of the error term.

OTHER ESTIMATION TECHNIQUES

- Bayesian linear regression applies the framework of Bayesian statistics to linear regression. In particular, the regression coefficients β are assumed to be random variables with a specified prior distribution. The prior distribution can bias the solutions for the regression coefficients, in a way similar to (but more general than) ridge regression or lasso regression. In addition, the Bayesian estimation process produces not a single point estimate for the "best" values of the regression coefficients but an entire posterior distribution, completely describing the uncertainty surrounding the quantity. This can be used to estimate the "best" coefficients using the mean, mode, median, any quantile, or any other function of the posterior distribution.
- Quantile regression focuses on the conditional quantiles of *y* given *X* rather than the conditional mean of *y* given *X*. Linear quantile regression models a particular conditional quantile, often the conditional median, as a linear function $â^T x$ of the predictors.
- Mixed models are widely used to analyze linear regression relationships involving dependent data when the dependencies have a known structure. Common applications of mixed models include analysis of data involving repeated measurements, such as longitudinal data, or

data obtained from cluster sampling. They are generally fit as parametric models, using maximum likelihood or Bayesian estimation. In the case where the errors are modeled as normal random variables, there is a close connection between mixed models and generalized least squares. Fixed effects estimation is an alternative approach to analyzing this type of data.

- Principal component regression (PCR) is used when the number of predictor variables is large, or when strong correlations exist among the predictor variables. This two-stage procedure first reduces the predictor variables using principal component analysis then uses the reduced variables in an OLS regression fit. While it often works well in practice, there is no general theoretical reason that the most informative linear function of the predictor variables should lie among the dominant principal components of the multivariate distribution of the predictor variables. The partial least squares regression is the extension of the PCR method which does not suffer from the mentioned deficiency.
- Least-angle regression is an estimation procedure for linear regression models that was developed to handle high-dimensional covariate vectors, potentially with more covariates than observations.
- The Theil–Sen estimator is a simple robust estimation technique that choose the slope of the fit line to be the median of the slopes of the lines through pairs of sample points. It has similar statistical efficiency properties to simple linear regression but is much less sensitive to outliers.
- Other robust estimation techniques, including the α-trimmed mean approach, and L-, M-, S-, and R-estimators have been introduced.

FURTHER DISCUSSION

In statistics, the problem of numerical methods for linear least squares is an important one because linear regression models are one of the most important types of model, both as formal statistical models and for exploration of data sets. The majority of statistical computer packages contain facilities for regression analysis that make use of linear least squares computations. Hence it is appropriate that considerable effort has been devoted to the task of ensuring that these computations are undertaken efficiently and with due regard to numerical precision. Individual statistical analyses are seldom undertaken in isolation, but rather are part of a sequence of investigatory steps. Some of the topics involved in considering numerical methods for linear least squares relate to this point. Thus important topics can be

- Computations where a number of similar, and often nested, models are considered for the same data set. That is, where models with the same dependent variable but different sets of independent variables are to be considered, for essentially the same set of data points.

- Computations for analyses that occur in a sequence, as the number of data points increases.
- Special considerations for very extensive data sets.

Fitting of linear models by least squares often, but not always, arises in the context of statistical analysis. It can therefore be important that considerations of computational efficiency for such problems extend to all of the auxiliary quantities required for such analyses, and are not restricted to the formal solution of the linear least squares problem.

Matrix calculations, like any others, are affected by rounding errors. An early summary of these effects, regarding the choice of computational methods for matrix inversion, was provided by Wilkinson.

APPLICATIONS OF LINEAR REGRESSION

Linear regression is widely used in biological, behavioral and social sciences to describe possible relationships between variables. It ranks as one of the most important tools used in these disciplines.

TREND LINE

A trend line represents a trend, the long-term movement in time series data after other components have been accounted for. It tells whether a particular data set (say GDP, oil prices or stock prices) have increased or decreased over the period of time. A trend line could simply be drawn by eye through a set of data points, but more properly their position and slope is calculated using statistical techniques like linear regression. Trend lines typically are straight lines, although some variations use higher degree polynomials depending on the degree of curvature desired in the line. Trend lines are sometimes used in business analytics to show changes in data over time. This has the advantage of being simple. Trend lines are often used to argue that a particular action or event (such as training, or an advertising campaign) caused observed changes at a point in time. This is a simple technique, and does not require a control group, experimental design, or a sophisticated analysis technique. However, it suffers from a lack of scientific validity in cases where other potential changes can affect the data.

EPIDEMIOLOGY

Early evidence relating tobacco smoking to mortality and morbidity came from observational studies employing regression analysis. In order to reduce spurious correlations when analyzing observational data, researchers usually include several variables in their regression models in addition to the variable of primary interest.

For example, suppose we have a regression model in which cigarette smoking is the independent variable of interest, and the dependent variable is lifespan measured in years. Researchers might include socio-economic status as an additional independent variable, to ensure that any observed effect of smoking on lifespan is not due to some effect of education or income. However, it is never possible to include all possible confounding variables in an empirical analysis. For example, a hypothetical gene might increase mortality and also cause people to smoke more. For this reason, randomized controlled trials are often able to generate more compelling evidence of causal relationships than can be obtained using regression analyses of observational data. When controlled experiments are not feasible, variants of regression analysis such as instrumental variables regression may be used to attempt to estimate causal relationships from observational data.

FINANCE

The capital asset pricing model uses linear regression as well as the concept of Beta for analyzing and quantifying the systematic risk of an investment. This comes directly from the Beta coefficient of the linear regression model that relates the return on the investment to the return on all risky assets.

ECONOMICS

Linear regression is the predominant empirical tool in economics. For example, it is used to predict consumption spending, fixed investment spending, inventory investment, purchases of a country's exports, spending on imports, the demand to hold liquid assets, labour demand, and labour supply.

ENVIRONMENTAL SCIENCE

Linear regression finds application in a wide range of environmental science applications. In Canada, the Environmental Effects Monitoring Programme uses statistical analyses on fish and benthic surveys to measure the effects of pulp mill or metal mine effluent on the aquatic ecosystem.

USES OF REGRESSION

Regression is a branch of statistical theory that is widely used in almost all the scientific disciplines. In economics it is the basic technique for measuring or estimating the relationship among economic variables that constitute the essence of economic theory and economic life. For example, if we know that two variables, price (X) and demand (Y), are closely related we can find out the most probable value of X for a given value of Y or the most probable value of Y for a given value of X. Similarly, if we know that the amount of tax and the rise in the price of a commodity are closely related, we find the study of regression is of considerable help to the economists and businessmen. The uses of regression are not confined to economics and business field only. Its applications are extended to almost all the natural, physical and social sciences.

The Regression attempts to accomplish the following:

1. With the help of regression coefficients we can calculate the correlation coefficient. The square of correlation coefficient (r), called coefficient of determination, measures the degree of association of correlation that

exists between the two variables. It assesses the proportion of variance in the dependent variable that has been accounted for by the regression equation. In general, the greater the value of r^2 the better is the fit and the more useful the regression equations as a predictive device.

2. A second goal of Regression is to obtain a measure of the error involved in using the regression line as a basis for estimation. For this purpose the standard error of estimate is calculated. This is a measure of the scatter or spread of the observed values of Y around the corresponding values estimated from the regression line. If the line fits the data closely, that is, if there is little scatter of the observations around the regression line, good estimates can be made of the Y variable. On the other hand, if there is a great deal of scatter of the observations around the fitted regression line, the line will not produce accurate estimates of the dependent variable.
3. Regression provides estimates of values of the dependent variable from values of the independent variable. The device used to accomplish this estimation procedure is the *regression line.* The regression line describes the average relationship existing between x and Y variables, *i.e.*, it displays mean values of X for given values of Y. The equation of this line, known as the regression equation, provides estimates of the dependent variable when values of the independent variable are inserted into the equation.

LIMITATIONS OF REGRESSION

In making estimate from a regression equation, it is important to remember that the assumption is being made that relationship has not changed since the regression equation was computed. Another point worth remembering is that the relationship shown by the scatter diagram may not be the same if the equation is extended beyond the values used in computing the equation. For example, there may be a close linear relationship between the yield of a crop and the amount of fertilizer applied. with the yield increasing as the amount of fertilzer is increased. It would not be logical, however, to extend this equation beyond the limits of the experiment for it is quite likely that if the amount of fertilizer were increased indefinitely, the yield would eventually decline as too much fertilizer was applied.

DIFFERENCE BETWEEN CORRELATION AND REGRESSION

Correlation and Regression are constructed under different assumption they furnish different types of information and it is not always clear as to which measure should be used in a given problem situation. The following are the points of difference between the two.

1. Correlation is merely a tool of ascertaining the degree of relationship between two variables and, therefore, we cannot say that one variable

is the cause and other the effect. For example, a high degree of correlation between price and demand for a certain commodity or a particular point of time may not suggest which is the cause and which is the effect. However, in regression thus making it possible to study the cause and effect relationship. It should be noted that the presence of association does not imply causation, but the existence of causation always implies association. Statistical evidence can only establish the presence or absence of association between variables whether causation exists or not depends purely on reasoning.

2. Whereas coefficient is a measure of degree of covariability between X and Y, the objective of Regression is to study the 'nature of relationship between the variables so that we may be able to predict the value of one on the basis of another. The closer the relationship between two variables, the greater the confidence that may be placed in the estimates.
3. In correlation analysis r_{xy} is a measure of direction and degree of linear relationship between two variables X and Y, r_{xy} and r_{yx} are symmetric ($r_{xy} = r_{yx}$), *i.e.*, it is immaterial which of X and Y is dependent variable and which is independent variable. In Regression the regression coefficients b_{xy} and b_{yx} are not symmetric, *i.e.*, $b_{xy} \neq b_{yx}$ and hence it definitely makes a difference as to which variable is dependent and which is independent.
4. There may be nonsense correlation between two variables which is purely due to chance and has no practical relevance such as increase in income and in crave in weight of a group of people. However, there is nothing like non sense regression

REGRESSION LINES

If we take the case of two variables X and Y, we shall have two regression lines as the regression of X on Y and the regression of Y on X. The regression line of Y on X gives the most probable values of Y for given values of X and the regression line of X on Y gives the most probable values of X for given values of Y. However, when there is either perfect positive or perfect negative correlation between the two variables ($r = \pm 1$) the regression lines will coincide, *i.e.*, we will have only one line. The farther the two regression lines from each other. The lesser is the degree of correlation and the nearer the two regression lines to each other, the higher is the degree of correlation. If the variables are independent, r is zero and the lines of regression are at right angles, *i.e.*, parallel to OX and OY.

It should be noted that the regression lines cut each other at the point of average of X and Y, *i.e.*, if from the point where both the regression lines cut each other a perpendicular is drawn on the X-axis, we will get the mean value of X and if from that point a horizontal line is drawn on the Y-axis, we will get the mean value of Y.

REGRESSION EQUATIONS

Regression equation, also known as estimating equations, are algebraic expression of the regression lines. Since there are two regression lines, there are two regression equations–the regression equation of X on Y is used to describe the variations in the values of X for given changes in Y and the regression equation of Y on X is used to describe the variation in the values of Y for given changes in X.

REGRESSION EQUATION OF Y ON X

The regression equation of Y on X is expressed as follows:

$Y = a + bX$

It may be noted that in this equation 'Y' is a dependent variable, *i.e.*, its value depends on X, 'X' is independent variable, *i.e.*, we can take a given value of X and compute the value of Y.

'a' is "Y intercept" because its value is the point at which the regression line crosses the Y-axis, that is, the vertical axis. 'b' is the "slope" of line. It represents change in Y variable for a unit change in X variable.

'a' and 'b' in the equation are called numberical constants because for any given straight line, their value does not change.

If the values of the constants 'a' and 'b' are obtained, the line is completely determined. But the question is how to obtain these values. The answer is provided by the method of Least Squares which states that the line should be drawn thought the plotted points in such a manner that the sum of the squares of the deviations of the actual Y values from the computed Y values is the least, or in other words, in order to obtain a line which fits the points best $\varepsilon(Y - Y_c)^2$, should be minimum. Such a line is known as the line of 'best fit'.

A straight line fitted by least squares has the following characteristics:

1. The straight line goes though the overall mean of the data $(\overline{X}. \overline{Y})$.
2. The deviations above the line equal those below the line, on the average. This means that the total of the positive and negative deviations is zero, or $\varepsilon(Y - Y_c) = 0$.
3. It gives the best fit to the data in the sense that it makes the sum of the squared deviations from the line, $\varepsilon(Y - Y_c)^2$, smaller than they would be from any other straight line. This property accounts for the name 'Least Squares'.
4. When the data represent a sample from a large population the least squares line is a 'best' estimate of the population regression line.

with a little algebra and differential calculus it can be shown that the following two equations. If solved simultaneously, will yield values of the parameters a and b such that the least squares requirement is fulfilled*:

$$\varepsilon Y - Na + b\varepsilon X$$
$$\varepsilon XY = a\varepsilon X + b\varepsilon X^2$$

These equitations are usually called the *normal equations.* In the equations εX, εXY, εX^2 indicate totals which are computed from the observed pairs of values of two variables X and Y to which the least squares estimating line is to be fitted and N is the number of observed pairs of values.

REGRESSION EQUATION OF X ON Y

The regression equation of X on Y is expressed as follows:

$$X_c = a + bY$$

To determine the values of a and b the following two normal equations are to be solved simultaneously:

$$\varepsilon X = Na + b\varepsilon Y$$
$$\varepsilon XY = a\varepsilon Y + b\varepsilon Y^2$$

NECESSARY NUMBER OF INDEPENDENT MEASUREMENTS

Consider a regression model which has three unknown parameters, $\hat{a}_0$, $\hat{a}_1$, and $\hat{a}_2$. Suppose an experimenter performs 10 measurements all at exactly the same value of independent variable vector X (which contains the independent variables X_1, X_2, and X_3). In this case, regression analysis fails to give a unique set of estimated values for the three unknown parameters; the experimenter did not provide enough information.

The best one can do is to estimate the average value and the standard deviation of the dependent variable *Y*. Similarly, measuring at two different values of X would give enough data for a regression with two unknowns, but not for three or more unknowns. If the experimenter had performed measurements at three different values of the independent variable vector X, then regression analysis would provide a unique set of estimates for the three unknown parameters in β.

In the case of general linear regression, the above statement is equivalent to the requirement that matrix XX is invertible.

STATISTICAL ASSUMPTIONS

When the number of measurements, *N*, is larger than the number of unknown parameters, *k*, and the measurement errors ε_i are normally distributed then *the excess of information* contained in (*N* - *k*) measurements is used to make statistical predictions about the unknown parameters. This excess of information is referred to as the degrees of freedom of the regression.

UNDERLYING ASSUMPTIONS

Classical assumptions for regression analysis include:

- The sample is representative of the population for the inference prediction.
- The error is a random variable with a mean of zero conditional on the explanatory variables.
- The independent variables are measured with no error. (Note: If this is not so, modeling may be done instead using errors-in-variables model techniques).

- The predictors are linearly independent, *i.e.*, it is not possible to express any predictor as a linear combination of the others. See Multicollinearity.
- The errors are uncorrelated, that is, the variance–covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.
- The variance of the error is constant across observations (homoscedasticity). (Note: If not, weighted least squares or other methods might instead be used).

These are sufficient conditions for the least-squares estimator to possess desirable properties, in particular, these assumptions imply that the parameter estimates will be unbiased, consistent, and efficient in the class of linear unbiased estimators.

It is important to note that actual data rarely satisfies the assumptions. That is, the method is used even though the assumptions are not true. Variation from the assumptions can sometimes be used as a measure of how far the model is from being useful. Many of these assumptions may be relaxed in more advanced treatments. Reports of statistical analyses usually include analyses of tests on the sample data and methodology for the fit and usefulness of the model.

Assumptions include the geometrical support of the variables. Independent and dependent variables often refer to values measured at point locations. There may be spatial trends and spatial autocorrelation in the variables that violates statistical assumptions of regression. Geographic weighted regression is one technique to deal with such data. Also, variables may include values aggregated by areas.

With aggregated data the Modifiable Areal Unit Problem can cause extreme variation in regression parameters. When analyzing data aggregated by political boundaries, postal codes or census areas results may be very different with a different choice of units.

REGRESSION DIAGNOSTICS

Once a regression model has been constructed, it may be important to confirm the goodness of fit of the model and the statistical significance of the estimated parameters. Commonly used checks of goodness of fit include the R-squared, analyses of the pattern of residuals and hypothesis testing. Statistical significance can be checked by an F-test of the overall fit, followed by t-tests of individual parameters.

Interpretations of these diagnostic tests rest heavily on the model assumptions. Although examination of the residuals can be used to invalidate a model, the results of a t-test or F-test are sometimes more difficult to interpret if the model's assumptions are violated. For example, if the error term does not have a normal distribution, in small samples the estimated parameters will not follow normal distributions and complicate inference. With relatively large samples, however, a central limit theorem can be invoked such that hypothesis testing may proceed using asymptotic approximations.

REGRESSION WITH "LIMITED DEPENDENT" VARIABLES

The response variable may be non-continuous ("limited" to lie on some subset of the real line). For binary (zero or one) variables, if analysis proceeds with least-squares linear regression, the model is called the linear probability model. Nonlinear models for binary dependent variables include the probit and logit model. The multivariate probit model is a standard method of estimating a joint relationship between several binary dependent variables and some independent variables.

For categorical variables with more than two values there is the multinomial logit. For ordinal variables with more than two values, there are the ordered logit and ordered probit models. Censored regression models may be used when the dependent variable is only sometimes observed, and Heckman correction type models may be used when the sample is not randomly selected from the population of interest. An alternative to such procedures is linear regression based on polychoric correlation (or polyserial correlations) between the categorical variables. Such procedures differ in the assumptions made about the distribution of the variables in the population. If the variable is positive with low values and represents the repetition of the occurrence of an event, then count models like the Poisson regression or the negative binomial model may be used instead.

INTERPOLATION AND EXTRAPOLATION

Regression models predict a value of the *Y* variable given known values of the *X* variables. Prediction *within* the range of values in the dataset used for model-fitting is known informally as interpolation. Prediction *outside* this range of the data is known as extrapolation. Performing extrapolation relies strongly on the regression assumptions. The further the extrapolation goes outside the data, the more room there is for the model to fail due to differences between the assumptions and the sample data or the true values. It is generally advised that when performing extrapolation, one should accompany the estimated value of the dependent variable with a prediction interval that represents the uncertainty. Such intervals tend to expand rapidly as the values of the independent variable(s) moved outside the range covered by the observed data.

For such reasons and others, some tend to say that it might be unwise to undertake extrapolation.

However, this does not cover the full set of modelling errors that may be being made: in particular, the assumption of a particular form for the relation between *Y* and *X*.

A properly conducted regression analysis will include an assessment of how well the assumed form is matched by the observed data, but it can only do so within the range of values of the independent variables actually available. This means that any extrapolation is particularly reliant on the assumptions being made about the structural form of the regression relationship.

Best-practice advice here is that a linear-in-variables and linear-in-parameters relationship should not be chosen simply for computational convenience, but that all available knowledge should be deployed in constructing a regression model. If this knowledge includes the fact that the dependent variable cannot go outside a certain range of values, this can be made use of in selecting the model — even if the observed dataset has no values particularly near such bounds. The implications of this step of choosing an appropriate functional form for the regression can be great when extrapolation is considered. At a minimum, it can ensure that any extrapolation arising from a fitted model is "realistic" (or in accord with what is known).

NONLINEAR REGRESSION

When the model function is not linear in the parameters, the sum of squares must be minimized by an iterative procedure. This introduces many complications which are summarized in Differences between linear and non-linear least squares

OTHER METHODS

Although the parameters of a regression model are usually estimated using the method of least squares, other methods which have been used include:

- Bayesian methods, *e.g.*, Bayesian linear regression
- Percentage regression, for situations where reducing *percentage* errors is deemed more appropriate.
- Least absolute deviations, which is more robust in the presence of outliers, leading to quantile regression
- Nonparametric regression, requires a large number of observations and is computationally intensive
- Distance metric learning, which is learned by the search of a meaningful distance metric in a given input space.

SOFTWARE

All major statistical software packages perform least squares regression analysis and inference. Simple linear regression and multiple regression using least squares can be done in some spreadsheet applications and on some calculators. While many statistical software packages can perform various types of nonparametric and robust regression, these methods are less standardized; different software packages implement different methods, and a method with a given name may be implemented differently in different packages. Specialized regression software has been developed for use in fields such as survey analysis and neuroimaging.

REGRESSION ANALYSIS

In statistics, regression analysis includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable given the independent variables — that is, the average value of the dependent variable when the independent variables are fixed. Less commonly, the focus is on a quantile, or other location parameter of the conditional distribution of the dependent variable given the independent variables. In all cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function, which can be described by a probability distribution.

Regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer causal relationships between the independent and dependent variables. However this can lead to illusions or false relationships, so caution is advisable: See correlation does not imply causation.

A large body of techniques for carrying out regression analysis has been developed. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be infinite-dimensional.

The performance of regression analysis methods in practice depends on the form of the data generating process, and how it relates to the regression approach being used. Since the true form of the data-generating process is generally not known, regression analysis often depends to some extent on making assumptions about this process. These assumptions are sometimes testable if a large amount of data is available. Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally. However, in many applications, especially with small effects or questions of causality based on observational data, regression methods give misleading results.

HISTORY

The earliest form of regression was the method of least squares, which was published by Legendre in 1805, and by Gauss in 1809. Legendre and Gauss both applied the method to the problem of determining, from astronomical observations, the orbits of bodies about the Sun (mostly comets, but also later the then newly discovered minor planets). Gauss published a further development of the theory of least squares in 1821, including a version of the Gauss–Markov theorem. The term "regression" was coined by Francis Galton in the nineteenth century to describe a biological phenomenon. The phenomenon was that the heights of descendants of tall ancestors tend to regress down towards a normal average (a phenomenon also known as regression towards the mean). For Galton, regression had only this biological meaning, but his work was later extended byUdny Yule and Karl Pearson to a more general statistical context. In the work of Yule and Pearson, the joint distribution of the response and explanatory variables is assumed to be Gaussian. This assumption was weakened by R.A. Fisher in his works of 1922 and 1925. Fisher assumed that the conditional distribution of the response variable is Gaussian, but the joint distribution need not be. In this respect, Fisher's assumption is closer to Gauss's formulation of 1821.

In the 1950s and 1960s, economists used electromechanical desk calculators to calculate regressions. Before 1970, it sometimes took up to 24 hours to receive the result from one regression. Regression methods continue to be an area of active research. In recent decades, new methods have been developed for robust regression, regression involving correlated responses such as time series and growth curves, regression in which the predictor or response variables are curves, images, graphs, or other complex data objects, regression methods accommodating various types of missing data, nonparametric regression, Bayesian methods for regression, regression in which the predictor variables are measured with error, regression with more predictor variables than observations, and causal inference with regression.

REGRESSION MODELS

Regression models involve the following variables:

- The unknown parameters, denoted as β, which may represent a scalar or a vector.
- The independent variables, X.
- The dependent variable, Y.

In various fields of application, different terminologies are used in place of dependent and independent variables.

Assume now that the vector of unknown parameters β is of length k. In order to perform a regression analysis the user must provide information about the dependent variable Y:

- If N data points of the form (Y,X) are observed, where $N < k$, most classical approaches to regression analysis cannot be performed: since the system of equations defining the regression model is underdetermined, there is not enough data to recover β.
- If exactly $N = k$ data points are observed, and the function f is linear, the equations $Y = f(X, \beta)$ can be solved exactly rather than approximately. This reduces to solving a set of N equations with N unknowns (the elements of β), which has a unique solution as long as the X are linearly independent. If f is nonlinear, a solution may not exist, or many solutions may exist.

- The most common situation is where $N > k$ data points are observed. In this case, there is enough information in the data to estimate a unique value for β that best fits the data in some sense, and the regression model when applied to the data can be viewed as an overdetermined system in β.

In the last case, the regression analysis provides the tools for:

1. Finding a solution for unknown parameters β that will, for example, minimize the distance between the measured and predicted values of the dependent variable *Y*.
2. Under certain statistical assumptions, the regression analysis uses the surplus of information to provide statistical information about the unknown parameters β and predicted values of the dependent variable *Y*.

MULTIPLE REGRESSION ANALYSIS

Multiple regression analysis is a powerful technique used for predicting the unknown value of a variable from the known value of two or more variables- also called the predictors. More precisely, multiple regression analysis helps us to predict the value of Y for given values of $X_1, X_2, \ldots, X_k$. For example the yield of rice per acre depends upon quality of seed, fertility of soil, fertilizer used, temperature, rainfall. If one is interested to study the joint affect of all these variables on rice yield, one can use this technique. An additional advantage of this technique is it also enables us to study the individual influence of these variables on yield.

By multiple regression, we mean models with just one dependent and two or more independent (exploratory) variables. The variable whose value is to be predicted is known as the dependent variable and the ones whose known values are used for prediction are known independent (exploratory) variables.

THE MULTIPLE REGRESSION MODEL

In general, the multiple regression equation of Y on $X_1, X_2, \ldots, X_k$ is given by:

$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + b_k X_k$

INTERPRETING REGRESSION COEFFICIENTS

Here b_0 is the intercept and $b_1, b_2, b_3, \ldots, b_k$ are analogous to the slope in linear regression equation and are also called regression coefficients. They can be interpreted the same way as slope. Thus if $b_i = 2.5$, it would indicates that Y will increase by 2.5 units if X_i increased by 1 unit. The appropriateness of the multiple regression model as a whole can be tested by the F-test in the ANOVA table. A significant F indicates a linear relationship between Y and at least one of the X's.

Once a multiple regression equation has been constructed, one can check how good it is (in terms of predictive ability) by examining the coefficient of determination (R^2). R^2 always lies between 0 and 1.

R^2-coefficient of determination

All software provides it whenever regression procedure is run. The closer R^2 is to 1, the better is the model and its prediction. A related question is whether the independent variables individually influence the dependent variable significantly. Statistically, it is equivalent to testing the null hypothesis that the relevant regression coefficient is zero. This can be done using t-test. If the t-test of a regression coefficient is significant, it indicates that the variable is in question influences Y significantly while controlling for other independent explanatory variables.

Multiple regression technique does not test whether data are linear. On the contrary, it proceeds by assuming that the relationship between the Y and each of X_i's is linear. Hence as a rule, it is prudent to always look at the scatter plots of (Y, X_i), i= 1, 2,…,k. If any plot suggests non linearity, one may use a suitable transformation to attain linearity. Another important assumption is non existence of multicollinearity-the independent variables are not related among themselves. At a very basic level, this can be tested by computing the correlation coefficient between each pair of independent variables. Other assumptions include those of homoscedasticity and normality.

Multiple regression analysis is used when one is interested in predicting a continuous dependent variable from a number of independent variables. If dependent variable is dichotomous, then logistic regression should be used.

DISCRIMINANT FUNCTION ANALYSIS

The main purpose of a discriminant function analysis is to predict group membership based on a linear combination of the interval variables. The procedure begins with a set of observations where both group membership and the values of the interval variables are known. The end result of the procedure is a model that allows prediction of group membership when only the interval variables are known. A second purpose of discriminant function analysis is an understanding of the data set, as a careful examination of the prediction model that results from the procedure can give insight into the relationship between group membership and the variables used to predict group membership.

For example, a graduate admissions committee might divide a set of past graduate students into two groups: students who finished the programme in five years or less and those who did not. Discriminant function analysis could be used to predict successful completion of the graduate programme based on GRE score and undergraduate grade point average. Examination of the prediction model might provide insights into how each predictor individually and in combination predicted completion or non-completion of a graduate programme.

Another example might predict whether patients recovered from a coma or not based on combinations of demographic and treatment variables. The predictor variables might include age, sex, general health, time between incident and arrival at hospital, various interventions, *etc.* In this case the creation of the prediction model would allow a medical practitioner to assess the chance of recovery based on observed variables. The prediction model might also give insight into how the variables interact in predicting recovery.

THE SIMPLEST CASE

The simplest case of discriminant function analysis is the prediction of dichotomous group membership based on a single variable. An example of the simplest case is the prediction of successful completion of a graduate programme based on the GRE verbal score. In this case, since the prediction model includes only a single variable, it gives little insight into how variables interact with each other in prediction.

With respect to the data file and purpose of analysis, this simplest case is identical to the case of linear regression with dichotomous dependent variables. Data of this type may be represented in any number of different forms: scatterplots, tables of means and standard deviations, and overlapping frequency polygons. Because overlapping frequency polygons have such an intuitive appeal, they will be used to describe how discriminant function analysis works.

PREDICTION ACCURACY

A single interval variable might discriminate between groups in an almost perfect fashion, not at all, or somewhere in between. For example, if one wished to differentiate adult males and females, one could collect information on how many bras the person owned, score on the last statistics test, and height. In the case of the number of bras, the discrimination would be very good, but not perfect (some women don't own any bras, some men do). In the case of the score on the last statistics test, little discrimination would be possible because males and females generally score about the same. In the case of height, some discrimination between adult males and females would be possible, but it would be far from perfect.

In general, the larger the difference between the means of the two groups relative to the within groups variability, the better the discrimination between the groups. The following programme allows the student to explore data sets with different degrees of discrimination ability.

MODELLING THE DATA

Discriminant function analysis is based on modelling the interval variable for each group with a normal curve. The mean of each group is used an estimate of mu for that group. Sigma for each group can be estimated by using weighted mean of the within group variances or using the standard deviation of that group. In the case of the weighted mean the variances are weighted by sample size and can be calculated either as the denominator for a nested t-test or as the square root of the Mean Squares Within Groups in an ANOVA, providing identical estimates. When using the weighted mean of the variances, one must assume that the generating function for each group produces numbers that in the long run have the same variability. In the simple case of dichotomous groups and a single predictor variable, it really does not make a great deal of difference in the complexity of the model if the variability of each groups is assumed to be equal or not. This is not true, however, when more groups and more predictor variables are added to the model. For that reason, the assumption of equality of within group variance is almost universal in discriminant function analysis.

The following programme allows the student to explore the relationship between different generating functions (poor, medium, or good discrimination; equal or unequal variances), sample size, and resulting model based on the sample. The student should verify that larger sample sizes provide resulting models that are more similar to the generating model. The student should also explore the effect of violations of the equality of variance assumptions on the resulting model.

PROBABILITIES OF GROUP MEMBERSHIP

The normal curve models of the predictor variables for each group and can be used to provide probability estimates of a particular score given membership in a particular group. In discriminant function analysis, the area in the tails under a normal curve model for a given group between points equally distant from mu is the probability of either point given that group. This probability is symbolized as P(D/G) on SPSS output.

For example, suppose that the normal curve model for a given group has a value of 13 for mu and 2 for sigma. The probability of a score of 16 would be the area in the tails of this normal curve between 10 and 16. The value of 10 was selected as the low score because 16 is three units above the mean (16-13 = 3) and 10 is three sigma units below the mean (13-3 = 10). This is all much easier visualized than stated.

PRIOR PROBABILITIES

Prior probabilities are the likelihood of belonging to a particular group given no information about the person is available. In the classical testing literature prior probabilities are called *base rates*. Prior probabilities influence our decisions about group membership. Prior probabilities will be symbolized as P(G). For example, $P(G_1)$ is the prior probability of a score belonging to group 1. Consider the case of prediction of completion of graduate school using GRE verbal scores. Suppose that 99% of the students who started the graduate programme successfully completed the programme (it was a really easy programme). Even if a student scored considerably lower than the mean of the successful group, programme completion would most likely be predicted because almost everyone finishes. Likewise, in a graduate programme where less than 10% of beginning students complete the programme, the likelihood of completion will be fairly low no matter how high the GRE verbal score.

3

The Activity of Learning Statistics

While activity may be viewed as an abstract construct, Leont'ev emphasises that in reality we always deal with specific activities, within a finite time, space and setting. His three levels of analysis provide three vantage points on a fundamental question for any investigation from the perspective of activity theory: the question of what an individual or group is doing in a particular setting.

In order to understand what a particular group of students is doing in learning Statistics, responses can be formulated in terms of each of these three levels. The first level concerns the cultural or historical milieu of the students' actions. This is an account of students' learning from the perspective of the global level of activity. To respond to the question at this level it is necessary to understand students' actions in the institutional context–while studying a university course.

A university has its own well defined social practices. Students' interpretations of these practices are bound up with their perceptions of what is expected of them and will be fundamental to their actions.

According to Wertsch:

- An activity setting is grounded in a set of assumptions about appropriate roles, goals and means used by the participants in that setting.

This setting guides the selection of actions and choice of tools. It also determines the function of the activity. Wertsch explains that settings are not determined by the physical context but are created by the participants in the activity. Further, assumptions about the setting are often implicit, rather than consciously identified. Participants may not identify what organises their performance. An individual's understanding of the setting emerges as a "byproduct" of interacting with others in it.. To say that a student is engaged in the activity of learning a university subject, Statistics, simply tells us that the student is working in a particular socioculturally defined setting. To understand activity at Leont'ev's second level of analysis one must look at the actions which, in his view, are defined by goals or partial goals.

In my view, goals are not automatic nor fixed in advance. They are tested by action and remain fluid throughout the process of selection and testing. This notion of goals as in flux - as being socially tested hypotheses - was expressed explicitly by one student whom I interviewed during her first year at university.

I've never really studied before this at all, and it's only been since last year that I actually applied myself to anything in academic terms. And so for me everything's just - what's the word... an experiment.

An experiment to see what works and what doesn't. Finally, a response to the question about a student's learning of Statistics can be formulated in terms of the tools and operations that the student has at her command. This is the third level of the activity of learning Statistics that must be investigated.

The availability of tools as mediating devices will partly be determined by the setting - for example what calculators are deemed appropriate for students to use. Tools and operations also depend on the student's personal history - her experiences, including the student's repertoire of skills, such as her ability to use a calculator or mathematical symbols. Affective elements are important as well.

If the student's memory of using a compass at school is linked to negative emotions, this tool will not be readily utilised. The tools or operations used by a student affect how successfully the student is able to carry out her intentions.

DEVELOPMENT AND PRINCIPLES OF ACTIVITY THEORY

HISTORICAL BACKGROUND

Many of Vygotsky’s conceptions greatly influenced and were further developed by his contemporaries, Luria, Leont’ev and other psychologists of the former Soviet Union. It is customary for Western authors to refer to Vygotsky, Luria and Leont’ev as a “troika” developing their theories in unity. Van der Veer and Valsiner dispel this unity as a myth, contending that different attitudes and thinking, as well as strong personal disagreements, marked the relationships among the three. Indeed Leont’ev did not co-author research publications with Vygotsky, although he did considerable empirical work in support of Vygotsky’s theories.

Leont’ev’s work on psychological functions began under Chelpanov at the Moscow Institute of Experimental Psychology, formerly the Psychological Institute. In 1923 Chelpanov was expelled from the Institute, which he had himself established in 1912, in favour of his one time student and strident critic, Kornilov, a reactologist. Leont'ev, under the guidance of Luria and Vygotsky, worked at the Institute under Kornilov, who was rushing to re-organise it along the lines of the new Marxist psychology.

Luria is cited as describing, with tongue in cheek, his early days under Kornilov as follows:

- Our Institute was supposed to reform the whole psychological science by abandoning Chelpanov's idealistic theory and creating a new materialist one.

Meanwhile the reform of psychology was proceeding in two forms:

1. First, by way of renaming things,
2. Second, by way of moving furniture.

In the early 1930's Leont'ev moved from Moscow to Khar'kov where he created his own school of psychology. His approach came to be known as the "activity approach in psychology" or the "psychological theory of activity". For Vygotsky and others of the cultural historical school, the focus of research was the development of mind and consciousness and how these are mediated by psychological instruments, especially words and other signs. To Leont'ev, in accordance with Marxism, the unit of psychological analysis is activity, and physical tools and concrete objects mediate the connection of humans and the natural world.

Activity theory is, however, grounded in and uses Vygotsky's concepts. For example Leont'ev re-interpreted Vygotsky's analysis of speech in terms of speech activity. Wertsch et al argue that the "action orientation" of Vygotsky has been obscured by his distinction between speech and language. To Vygotsky, as to his Russian colleague Bakhtin, a linguistic philosopher, speech is a process, language is a semiotic means. That is, speech is a form of action. Language, while "it has the power to shape speaking and thinking" is not an action in Vygotsky's terms.

There is a difference between speech, as utterances, and language, as a system of communication. This difference is in the emphasis on different aspects–"the vocal and the semantic" which Vygotsky saw as related, but with different patterns of development. Vygotsky tended to focus on speech as opposed to language. However, he also paid attention to the symbolic properties of the mediator as central to the action of meaning making. Leont'ev, however, stressed the external active form of speech. While Vygotsky regarded word meaning as the appropriate unit of analysis, Leont'ev based his analysis on activity. In both cases these units are cultural. The title of Vygotsky's last volume *Thinking and Speech* highlights the "action dynamic" in Vygotsky's thesis. Indeed, the important part played by Vygotsky in the creation and development of the activity paradigm itself was acknowledged by Leont'ev in his obituary of Vygotsky.

Leont'ev's interpretation of Vygotsky's theories must be seen in the light of his attempts to rehabilitate Vygotsky's work after the latter's work was banned by the Russian authorities soon after his death in 1934. In the repressive climate of the former Soviet Union in the pre-Glasnost era, researchers were forced to demonstrate their loyalty to the prevailing ideology. Vygotsky came under increasing attack and his refusal to adhere to strictly Marxist principles regarding the primacy of material production was cited by Petr Zinchenko and other contemporaries of Vygotsky as a fundamental error.

Vladimir Zinchenko, the son of Petr Zinchenko, affirms that Vygotsky and any who supported him were in danger of losing their freedom, if not their lives. He notes that cultural-historical psychology emerged during the "dismemberment" of Russian culture, en route to communism, while the background to activity theory was the "unprecedented" enslavement of the country's peasants. Hence, the very emergence of these theories, was, in V. P. Zinchenko's opinion, miraculous. Bodrova and Leong applaud the courage of those like Leont'ev who continued to expand and elaborate on Vygotsky's theories after the latter's death, despite the risks.

In their opinion, such people were then able to revive Vygotsky's ideas in the late fifties after the bans had been lifted. Other commentators paint a less rosy picture of the actions of Leont'ev and Luria, indicating that they did not always stand up for academic freedoms. Clearly Russian society at the time was a maelstrom and simplistic judgements cannot be made about those trying to steer their courses through the storms.

In V. P. Zinchenko's view, Vygotsky's students and colleagues, such as Leont'ev, did not simply camouflage their earlier views after Vygotsky's death but worked in a new direction.

Leont'ev's line of research was distinct from Vygotsky's and:

- *Today we are dealing with two scientific paradigms:* – Cultural-historical psychology
  - The psychological theory of activity.

These two strands of research can be seen as mutually amplifying and enriching each other.

PRINCIPLES OF ACTIVITY THEORY

Davydov an "avowed adherent" of activity theory, and arguably one of its most well known Russian advocates in the contemporary Western world, explains that Marx took activity as a key concept to explain the development of economic life. This concept of activity was worked out in the context of German classical philosophy. Philosophers, such as Kant and Hegel formulated the groundwork for the concept of activity, its structure and its formation in the sociocultural realm. Marx emphasised the manufacture of artefacts as satisfying the basic needs of humans.

In his view, production, or labour, is the means by which man changes nature and at the same time, his own nature. This production depends on the use of tools. Mental abilities, too, arise in the process of material production and the inter-personal relationships surrounding it. According to Marx, human activity involving labour is the foundation for collective and individual consciousness.

That is, activity is the basis of “the personal and social identity of the individual”.

In Russia, after the October 1917 Revolution, the concept of activity was developed in terms of disciplines such as philosophy and psychology. There were many different versions of activity theory of which Leont’ev’s became the most well known. Indeed, Gilgen and Gilgen comment that in scholarly publications written in English as well as in Russian, Leont’ev’s account of Soviet psychology “predominated from the 1950’s to the end of the Soviet regime in 1991". In part this is due to the prominent positions held by Leont’ev in Russia, which included Chairman of the Department of Psychology, Moscow State University. Lektorsky, a Russian philosopher, is cited by Cobb et al as an expert in activity theory.

Lektorsky explains cognition in terms of Marxist philosophy as follows:

- Marxist philosophy asserts that cognition is founded on practical activity and that the latter must be understood in its specifically human characteristics, to wit, as collective or joint activity, in which the individual enters upon definite relations with other persons, as mediated activity in which man places between himself and an external naturally emerging object other man made objects functioning as the implements of activity: and finally, as a historically developing activity carrying in itself its own history.

In this, somewhat dense, sentence are the basic tenets of activity theory. The first of these is that activity is the explanatory principle in psychological theory. Secondly, Lektorsky equates cognition as socialisation, as Vygotsky claimed, with cognition as the “interiorization” of external actions, in Leont’ev’s terms by describing all practical activity or labour as being collective. That is, all activity is of a social nature. Thirdly, activity is mediated by cultural tools which shape the individual’s functioning. Finally there is the idea of the mutual formation of mind and the social world in history. This last idea is expressed cogently by Varela et al who propose that mind and the world co-emerge by structural coupling, much as the colours of flowers and the vision of bees co-evolve. Davydov makes the point that the term “activity” does not have the same frame of reference everywhere, in research in the social sciences. In English, the word covers all kinds of overt and covert actions.

In Leont’ev’s activity theory the term:

- Refers exclusively to events in which actions are judged by their intention to change reality.

That is, activity is differentiated from other human actions by means of being goal driven.

Davydov defines activity in this framework in the following way:

- Activity is a specific species of human societal existence which aims at a goal directed change of physical and social reality.

So, individual activity is embedded in and transforms the social and cultural life of the community. At the level of the individual, activity brings together a goal, the means for realising this goal, the actual process of transformation and its effect or outcome. Each of these is organised within a societal setting. The goal of activity emerges first as a schematic model of the desired outcome of the actions. That is, practical actions start with mental plans. This has the effect of creating opportunities–of rising the limits as they exist and so increasing the options available.

The "system" of collective activity which relates people to each other, their cultural history and their realm of action is called an "activity system" by Engeström. Taken as a unit of analysis, it confers meaning on seemingly random and individual behaviour. The activity system of people acting in their historically formed and social world is in a process of change and contradiction.

The dialectics of the fusion of the subject and the external world are explained by Lektorsky by quoting Lenin as follows:

- The reflection of nature in man's thought must be understood not 'lifelessly', not 'abstractly,' *not devoid of movement, not without contradictions,* but in the eternal *process* of movement, the arising of contradictions and their solution.

It is these perturbations which lead to development of the system including the development of its participants. The following parts clarify the main principles underlying activity theory.

Activity is the Explanatory Principle of Psycho Logy

Leont'ev, stressed that consciousness is "generated" by and inseparable from external activity. In his view, based on Marxism, activity must have associated external behaviour and mental images are direct products of contact with the world of objects. In this thinking, Leont'ev was influenced by others including, importantly, Rubinshtein and Vygotsky but his thinking differed from each of theirs.

Rubinshtein, like Leont'ev, proposed that consciousness and behaviour represent a unity. Rubinshtein defined individual consciousness as: the person's ability to regulate not only movements, acts and operations taking place at a particular time but also the entire life span. Rubinshtein's ideas accord with those of Piaget, in that he focused on the roles of individuals' independent explorations and interactions with the objective world as the sources of their reflections of it.

That is, Rubinshtein's position puts the subject at the centre of psychology. Vygotsky put interaction and communication among people at the centre of individual psychological development. As discussed earlier in this stage Vygotsky explained the development of higher mental faculties in terms of historically developed social interactions. In his theory, it is these internalised social interactions and communications that are the basis of cognitive development.

Diverging from both Vygotsky and Rubinshtein, Leont'ev proposed that activity is at the centre of a person's connection with others and the external world. Activity includes interaction with others and it also includes reflection, but Leont'ev's emphasis is on external actions. That is, according to Leont'ev individual psychological experiences are expressed through external, practical activity. Activity connects individuals to their communities and connects material objects to the psychological faculties that illuminate these objects.

This leads to the question of how objects are connected with the mind. Bakhurst, an American interpreter of activity theory, solves the problem by referring to the meaning of objects acquired by them in social practice. That is the "ideal" existence of objects is completely different from its physical nature - the stuff of which it is made. It is an artefact endowed with social significance. To illustrate this, Bakhurst describes a pen as a form which objectifies a social purpose.

The meanings that a community give to an object transform it into a cultural artefact which has ideal properties. These properties become accessible to our minds when we participate in the social practices for which the object was developed. Hence, writing with a pen gives it its"reality". This idea of objects as mediators of sociohistorical experience is also applicable to symbols, referred to by Lektorsky as symbols-in-practice, rather than just symbols. For example, Lektorsky argues that when we read text and try to understand it, we focus on the content expressed with the help of letters, but ignore any characteristics of the letters that are "indifferent" to meaning making, such as their individual shapes.

Activity is Purposive

According to Leont'ev:

- There can be no activity without a motive. 'Unmotivated' activity is not activity devoid of a motive: It is activity with a motive that is subjectively and objectively concealed. Hence to Leont'ev, some motive, whether conscious or unconscious, always stands behind activity. Motives, needs and desires are translated into action by means of goals.

Motivation and goal formulation are inherently social. They connect psychological phenomena to the social world. Leont'ev proposes that when a desire meets the object of desire, it is filled"with content from the surrounding world". That is, individual actions are embedded in larger systems of activity called"socially patterned systems of activities" by Scribner and are directed towards satisfying motives. These systems of activity are socially patterned in three ways: the goals they satisfy are socially generated, they are carried out through social means and they are connected to a system of collective activity which has developed over human history. The idea of a goal follows from the Marxist concept of labour.

This notion is interpreted by Davydov as follows:

- At the end of the labour process a result is obtained which existed in the person's conception even at the start of the process-that is, ideally.

That is, according to the theories of Marx and Engels, the human ability to formulate and carry out conscious goals provides the cognitive framework for humans to enter into production. In line with the Marxist paradigm, Leont'ev's theory postulates that human motives are determined by the division of labour in society, and that within this framework, specific, concrete actions are related to practical goals. Leont'ev's theory, the idea of individual potential emerges from his postulate that actions are consciously directed at a goal.

The intention of cognition, in the sense of directedness, implies an awareness of what possibilities for action there are, and an understanding of how the context constrains the actions or enables the fulfilment of these possibilities. Historically, goal formation was one of the major research topics in psychology in the former Soviet Union.

Current research in the West has also provided empirical support for the importance of personal goals as directing the learning of university students. The research highlights the importance of goals in the self regulation of students' activities, the complexity of the dynamics of goal formation and development in academic settings, and the ways goals are responsive or resistant to the environment.

In contrast to Leont'ev's ideas, however, this research gives an account of students' goals that:

- Precludes any easy invocation of goals as causes of particular sets of actions.

That is, goal directedness is neither transparent nor unequivocal.

Activity is Mediated

Activity can be thought of as having a tripartite structure consisting of the individual, the environment and the tool, which, as a cultural mediator, is both the agent and reflector of change. Leont'ev explains that the tool mediates activity and in this way connects humans with concrete objects and also with other people. He specifies that it is not the cultural tool, as an artefact, but the use of that tool that mediates or shapes mental processes.

That is:

- The formation of uniquely human functional systems takes place as a result of mastering tools and operations.

Moreover, Leont'ev contended that such mediators do not simply facilitate human thinking and agency. As the social structure and tools of production change, so does human thinking. This relates to Vygotsky's idea that as technical tools alter the operations of labour, so psychological tools transform "the entire flow and structure of mental functions".

This premise is most pertinent to our society with its rapid technological transformations. For example, the ways we use our memories and other mental abilities have changed dramatically with "extra-cerebral" tools like the Internet, and even calculators. These have taken over the function of storing vast amounts of information and of processing information by means of complex algorithms, thus allowing for increased expertise and more complex mental actions, such as problem solving.

Activity Theory Emphasises the Interdependence and Interaction of Sociocultural and Cognitive Processes

The view that cognitive abilities of the child or novice emerge from interaction with adults or experts was developed by Vygotsky who proposed that cognitive processes are carried out interpsychologically then intrapsychologically, or from the social to the individual plane. That is, they are first carried out initially with help from an adult or expert, then transformed to a self directed carrying out of the task, via the zone of proximal development. Speech is the major mediational means which directs this process.

In Leont'ev's modification of this theme, activity is the major source of cognitive development. Psychological functions, too, such as thinking, are actions in Leont'ev's terms. Leont'ev proposed that cognition can exist only in the presence of and embedded in social consciousness - what Leont'ev, echoing Vygotsky, called"co-knowing".

It lives only in an environment of socially constituted meanings and is developed by the appropriate use of cultural tools. Leont'ev proposed that inner mental processes originate from external activity which serves a practical, socially developed purpose.

That is, thought is"not in the head", in a private world of consciousness, but in the performance of social activities. External practical activity is internalised, that is, changed into processes that take place at the level of consciousness. Leont'ev expressed this as"... life underlies consciousness". During the transition these external processes become generalised, verbalised, abbreviated and become the means for development beyond the bounds of only external activity. Leont'ev quotes Piaget as formulating that this conversion"leads from the sensory-motor plane to thought".

A specific account of this essentially social process of transformation is provided by Davydov and Andronov. They call it a"test tube" example of the emergence of mathematical thinking. In their study of young children working on addition problems, the mental act of addition initially takes place as an overt action, a sweeping hand movement as the number is pronounced. The gesture, in turn, becoming a symbol for the number, is reduced, and, together with speech, becomes the means for organising the child's mathematical thinking during the problem solving.

Davydov explains that through instruction:

- Every individual appropriates to himself, converts into Forms of his own activity, the means and methods of thought that have been created by society at that historical epoch.

This idea stresses the sociohistorical roots of cognition. It also emphasises the engagement of the individual. The term "appropriate", like the term "construct" highlights the active role of the learner. This is particularly relevant to students learning Statistics.

Statistics is an artefact of our technological era with a history of development in society and it is perceived differently by each individual as he or she "appropriates" it, in Davydov's term.

Cobb et al sums up the different senses in which Vygotsky and Leont'ev view an individual's actions as being social as follows.

- Vygotsky views personal actions as involving direct interactions with others.
- Activity theorists view individual actions in terms of the broader sociocultural system.

The latter view is a stronger, more encompassing account of social.

Minick points out, as a key element of the activity framework, that:

- Psychological characteristics are conceptualized not as characteristics of the individual but as characteristics of the individual-in-(social)-action.

This stronger meaning of social has been applied in many studies of situated mathematical activity such as investigations into the delivery of milk crates or into supermarket shopping. In these studies, and in my own investigation, individuals participate in social activity which may, but does not always, include face to face interaction.

Lektorsky argues that there is a continuity between Vygotsky's notion of social activity as the determinant of psychological development and Leont'ev's focus on mental abilities being derived from labour or practical activity. According to Lektorsky, a person's cognition is shaped by interacting with others who teach him to use man-made things. Lektorsky combines Vygotsky's and Leont'ev's proposed mechanisms for the development of the psyche, by explaining that an individual's consciousness is founded in the direct links between practical activity, cognition and "the living communicative connections with other persons".

ILLUSTRATIONS OF THE FEATURES OF ACTIVITY THEORY: EXAMPLES FROM MY PRACTICE

So far, I have outlined a model for understanding learning suggested by the theory of activity. In summary, in explaining cognition, the emphases of activity theory are on its connected elements and mediated structure, how it functions by means of goal directed actions and on its cultural and historical development.

Engeström points out that activity theory does not offer ready made techniques and procedures.

Its conceptual tools and methodological principles:

- Have to be concretized according to the specific nature of the object under scrutiny.

In this part I present some of the elements of the theoretical framework which are of interest to me in defining the activity of learning Statistics. I shall attempt to illustrate these factors with examples drawn from my teaching practice. The part concludes with some implications of the activity theory approach for understanding learning and teaching.

DIALECTIC OF STUDENT, ACTIVITY AND SETTING

My perspective stresses that it is not fixed or pre-ordained features of the setting or of the individual learner that determine the learning. The student as a learner and the context of her learning are mutually constituted in social action.

I agree with Ratner's interpretation that:

- Social activity not only determines the content of psychological phenomena. Activity also conditions the particular fields of life in which a psychological phenomenon is employed.

Ratner points out that culture means that there is a specific, practical system operating. This determines the distribution of power, privileges and resources. It establishes the ways in which tasks are separated from one another, for example, school learning and vocational training. It influences how people act and relate to each other in different social practices, such as during work and at home. Ratner argues that, although culture is commonly recognised as the sharing of conceptual meanings and understandings, it also includes these practical aspects.

These must be recognised for the important ways in which they affect people's psychological processes. Hofstede describes different dimensions of culture, such as gender roles, individualism versus collectivism and distribution of power, that are important in the organisation of people's behaviour and thinking.

In Australia, higher education is undergoing massive changes, in part due to political and economic policies.

These policies are transforming the ways in which educational processes are framed. For example, the ways that educators examine students, their priorities with respect to teaching and research, the efforts they make to reform the curriculum, the demands on their time and the resources available to them reflect the political and economic environment. These policies even transform perceptions of the past. For example, current "cut backs" in departmental budgets instantaneously transform the past into a golden era–a supposed time of plenty from which things have been cut back.

The educational context shapes the ways in which students learn. In turn, students' actions change the teaching situation which surrounds them, including the actions of educators. For example, if an educator teaches the same topic for two or more years, the teaching strategies and methods of communication which "worked" one year may be dramatically changed the following year. Such changes may reflect the social environment produced by the members of each new class.

The activity theory approach to education asserts that knowledge, and ways of gaining knowledge, have a social ontogeny. One implication of this is that students' previous experiences can be ameliorated. This fits with my own experience of teaching situations. For example, it was the case for Jane, an older student whom I helped with Statistics at the Mathematics Learning Centre. She felt that she had not succeeded with mathematical subjects before coming to university. She did not believe herself to be capable of doing so.

I interviewed her after she had achieved very good results in Statistics and she described her feelings as follows:

- When I've got my notes and my calculator I feel invincible.... I found that I actually knew more than some other people. Instead of being the one who knew the least. And that gave me an enormous sense of power - or a bit of power anyway - that I could actually know more than somebody else.

As in Jane's case, not only do students' perspectives and insights into a particular field of knowledge change, but also their very self concepts. Slotnick, Pelton, Fuller and Tabor propose that in answering a student's questions, an instructor is not only helping that person advance her skills in reasoning and logic and providing a role model for dealing with complex issues, but is also "validating" that person, that is, confirming her identity.

Changes in students have been empirically demonstrated in the case of mature students who attend university. These transformations are related to their experiences and to their goals, which unfold and are structured by the social and cultural meaning of the events. Further, this social meaning is embedded in specific, practical situations of activity. Jane's activity took place in a particular university within the wider Australian culture, surrounded by the customs, norms and gender roles pertaining both to the institution and to our society. Each of these arenas - her immediate social environment, the institutional setting and the wider Australian society - are contextual orbits within which Jane's activity was organised.

By the word "orbits" I am emphasising that the structures operating on Jane were undergoing changes themselves, while seeming to be static - as the earth, surrounding us, appears static although we know it is moving. It is these contexts that influenced the way Jane thought about herself and structured the tasks in which she was engaged. They determined which of her actions were rewarded and how they were rewarded and what demands were placed on her.

VIGNETTES FROM PRACTICE

The theoretical approach offers ways of interpreting students' actions. In order to clarify the ideas I have introduced in this stage, I summarise my interpretation of key features of activity theory by means of examples. The following excerpts are from transcribed interviews with students whom I assisted with various statistics courses. These are snapshots, taken from different angles, of an activity system of students acting in their institutional worlds. They capture images of ongoing and dynamic processes.

Activities are Distinguished by their Energising Motives

Brenda, first year statistics student, age 23, said this:

- Like to understand what I'm doing. I don't just want to pass statistics, scrape through. I want to work. My whole attitude's totally different from when I was at school. I didn't care when I was at school. I became a Christian a couple of years ago. And so my whole attitude changed and now I guess I've got a goal and now I've got a reason for being alive basically. I never did well at anything at school and last year the course I did, out of 160 people I came first. That's amazing. That for me is amazing.

Brenda's quote illustrates the essence of activity as encompassing her total perspective on life. Her actions in the educational setting were guided and directed by her underlying raison d' être–in this case spiritual inspiration, and this resulted in a very favourable outcome.

Activity Theory Emphasises the Interdependence and Interaction of Learning Processes and Cultural Background

The following excerpt is from an interview with a first year statistics student, Huong, who recently entered Australia. Huong expressed a mismatch between her schooling in Vietnam and the ways she was expected to learn statistics in Australia. There is a big difference between the math I did at high school and this statistics course. Everything is new for me. From the beginning of this year I learn how to use a calculator and how to work with the terms in math, how mathematics terms mean in English. But it is hard to learn to solve problems because in my country the way to learn is just to memorise by heart. We had no chance to practice to think for ourself. Just to memorise what the teacher said to us.

From the activity theory point of view, this excerpt is an invitation to explore the roles of memorisation in different cultures. For example, Biggs cites studies investigating how Asian and Australian students use memorising. These show, that in our culture, memorising with the intention of reproducing material may not lead to understanding, and is then inappropriate to the task of learning. For Chinese students in Hong Kong, it is more likely that memorising leads to understanding–the way that material is internalised is part of a deep approach to learning. Further exploration on the relation between memorisation and understanding for Chinese learners is provided by Marton et al.

Actions are Socially Formulated

Eve: Second year Psychology student describing her mainstream Statistics tutorials:

- Its much more like you imagine university maths to be nobody will ask questions. It doesn't seem to be the done thing.

Hence Eve's actions were prescribed by unwritten codes of social conduct in tutorials, as she perceived them.

Activity Involves Self Regulation

Student repeating Psychology II course:

- I can’t use any of my own knowledge for Statistics. I can’t use any of my analytical or philosophical skills or just like experiences in life– it’s rote learning and its just so different.

This quote expresses the mutual structuring of this student’s conceptions of learning Statistics and her regulation of her actions to learn it.

IMPLICATIONS FOR THE ANALYSIS OF LEARNING

The vignettes illustrate some aspects of my perspective drawing on activity theory. This perspective has implications for the ways in which we view the learning process. These implications are not, of course, unique to this paradigm.

Indeed, viable research and teaching may follow from more than one paradigm.

- The activity framework implies a commitment to investigating the learning process as a dynamic process and one in which society and the individual interact and evolve rather than the student as a closed system or passive receiver of ready made information.
- The theory presents different levels at which learning can be analysed: the global level of the context of the learning, the level of the goal directed actions, that is, how the task was carried out and the level of the operations, which, according to Leont'ev, depend on the means by which an action is carried out.
- It emphasises the social dimension of learning: certain behaviour is acceptable in an arena of action, such as a tutorial session, other behaviour is not.
- It sensitises us to cultural effects. This is particularly pertinent for studying the learning of statistics as our current Western consumer society is immersed in statistics yet not all students are familiar or comfortable with this.
- It takes account of goals. The degree to which the goals set by the educator fit with those of the students will affect their involvement in the tasks set for them.
- A Vygotskian perspective recognises the importance of language and symbols. Statistics is a branch of mathematics which has a large verbal component. The language used in statistics, described by one European born student as"English acrobatics" is often the very aspect that confuses students.
- A further strength of the Vygotskian approach is its emphasis on underlying metacognitive skills, such as self regulation. Empirical research shows that these are important for the development of appropriate learning strategies.
- In sharp contrast to psychological theories which view cognitive processes as separate from the emotional domain, activity theory emphasises the

inseparability of students' feelings from their thinking. These cannot be considered as discrete, additive components. Leont'ev concurred with Vygotsky that motivation and the affective volitional sphere were the"hidden" plane of thinking.

According to Vygotsky, the separation of intellect and affect as subjects of study:

- Makes the thought process appear as an autonomous flow of 'thoughts thinking themselves,' segregated from the fullness of life, from the personal needs and interests, the inclinations and impulses, of the thinker. Rather than needs, feelings and emotions being treated as residing within the individual, influenced by the environment - a precondition of action, an activity theory approach regards these as unified aspects of human functioning.

ISSUES CONCERNING CONTEMPORARY INTERPRETATIONS OF ACTIVITY THEORY

PROBLEMS IN THE THEORETICAL ASSERTIONS OF ACTIVITY THEORY

A number of unsolved problems and issues pertaining to activity theory have been highlighted in recent research. These include the meaning of the construct "activity"; issues concerning learning by "internalization"; the nature of knowledge and of reality and methodological problems concerning activity as a unit of analysis. Davydov presents a number of problems pertaining to "activity", itself. Firstly, despite his own extensive work clarifying it, he regards the basic meaning of activity to be inadequately conceptualised. Activity, in his opinion, is insufficiently differentiated from other human actions, despite the emphasis on its goal orientation. Davydov argues that each type of activity, such as art, politics or labour, must be understood in terms of its own cultural historical formation. For example, what types of instructional activity were privileged in specific historical periods? He also holds that each type of activity is embedded in processes of interpersonal relationships. These, too, must be explicated in order to understand the activity. In my project the activity of learning Statistics is interpreted in an institutional setting, which is culturally and historically formed and organised. It relates also to the interactions among students and academics.

One of the most widely criticised aspects of activity theory concerns the process of internalisation. This is a notion of Vygotsky which is also central to Leont'ev's theory. Two major concerns are highlighted by Bauersfeld. One pertains to how individual activity is formed on the basis of collective activity. The second aspect relates to the idea that what is internalised is ready made and singular.

The first issue was raised by Davydov who criticised Leont'ev for pointing out the similarities, but not the differences in structure, between collective activity and individual activity. To Davydov it is not clear how individual functioning, such as remembering, would be derived from a differently structured collective activity. One explanation that I offer in the case of remembering is that collective memories are individually accessed as metaphors. For example, children are told fairy stories about dangers which have their roots in "tribal" memories.

The chances of an Australian child encountering a wolf on route to her grandmother are remote for the vast majority. However, this "tribal" memory serves a useful purpose in the child's current behaviour, assisting her to cope with life by warning her of its dangers and also enabling her to access and express her emotions. Also, individual memories can be jogged and even amended by means of external aids such as photographs or the recollections of others. The second concern relates to the idea that what is internalised is ready made and singular. This seems to suggest that there is a historically formed, objective reality "out there", which can be used to lead development. This aspect leads to problems explaining creativity and also suggests that knowledge is ready made.

To Rogoff the idea of internalisation suggests that:

- Something static is taken across a boundary from the external to the internal.

Rogoff like Davydov uses the term "appropriate" rather than "internalise" to emphasise the person's participation in an activity. Rogoff also proposes that "participatory appropriation" is a "process of transformation", whereby the boundary itself is questioned, since a person who is participating in an activity is part of that activity.

Rogoff comments that Vygotsky's notion of internalisation may be similar to her idea of appropriation in that transformation is inherent to his notion. To Leont'ev, too, internalisation is a dynamic, active process. However, Rogoff's term "participatory appropriation", in which the person and the social world are not separate, contrasts with the "internalisation" of an event or knowledge which is external to the individual.

The idea of historically formed, crystallised external knowledge can lead to a positivist and reductionist view of mathematics. Teaching mathematics from this perspective would mean guiding students to the "truth", using direct instruction, powerful mediational means and explication of reflective strategies. In contrast, teaching mathematics is problematic from the constructivist perspective. A teacher cannot transmit knowledge or expose information.

Burton emphasises the:

- Interaction between individuals, society and knowledge out of which mathematical meaning is created.

This is a non-positivist position which fits with the metaphors of co-emergence of person and knowledge and of the interactionist perspective. My understanding agrees with the latter positions. Rather than mathematical knowledge as non-negotiable and absolute "truth", it is constituted in the relationships among what is learned, those who learn or use it and the society within which it develops.

On a methodological level, Leont'ev has been criticised as viewing activity both as a unit of life which can be analysed by means of different levels and a means of investigating life itself. Kozulin considers this to be a consequence of the rejection by Leont'ev and his followers of Vygotsky's notions concerning the role of semiotic systems, or cultural signs in general.

In his effort to adhere to Marxist principles, Leont'ev saw activity as providing not only the subject matter but also the methodology or explanatory principle for psychology. To Vygotsky, in contrast, higher mental functioning cannot be explained without reference to semiotic mediation.

To avoid this circularity of explaining activity by means of activity, I will treat students' reported activities in learning Statistics as the object of my investigation. Actions and operations will be regarded as components of their activity. Moreover, in my investigation I will confine my focus to what is accessible through students' awareness. Leont'ev insists on all actions as goal directed. However, we do not always articulate, and are not necessarily even aware of, the motives for our actions. Activities that comprise of purposeful actions are, however, accessible to research through the awareness of the participants or actors.

TAKING ACTIVITY THEORY INTO THE FUTURE

The ideas of Vygotsky, Leont'ev and other theorists of the former Soviet Union have provided and continue to provide fruitful areas of research both in the East and in the West. Lektorsky comments that in the birthplace of activity theory, there is much criticism, in intellectual circles, of activity theory. Firstly, there is the understandable repudiation by these people of totalitarianism and philosophies based upon it.

As Davydov concurs, activity theory:

- Came to be fiercely criticised because it was connected with the idea of technological progress in which the ethical dimension is put between brackets.

Secondly, and linked to the previous critique, it is argued that the emphasis in activity theory on objectivity, as opposed to subjectivity, has led to the suppression of ideas about the creativity of individuals. That is, activity theory is criticised as representing humans not as creative beings but as interiorisers of ready made standards of behaviour and cognition.

While concurring with these criticisms, Lektorsky argues that Vygotsky's and Leont'ev's ideas should not all be rejected with Marxism. He also advocates that activity theory should be developed in new and fruitful ways. He argues that to understand activity we would do well to return to the ideas of Vygotsky who stressed not only interiorisation but exteriorisation–the ways people create new standards and rules. In this view, human beings are essentially creative beings who "determine themselves through objects that they create".

Davydov argues that the concept of activity refers primarily to the:

- Openness and universality of human actions. Consequently, 'activity' should be interpreted as a form of creativity which is determined by the actual cultural-historical context.

To Davydov totalitarianism intruded on a concept of activity “torn loose” from its own cultural-historical roots in the work of Vygotsky. Davydov advocates the development of a multi-disciplinary activity theory, one that combines the differences of emphases of the semiotically oriented researchers with those taking as their paradigm Leont’ev’s conceptualisation of the activity aspect of human existence. Indeed, in Davydov’s opinion, a “broad spectrum” approach must be taken to solve the problems. That is, analysis must take place from a multi-disciplinary point of view.

Only in this way, he feels, can we:

- Shape our inheritance from the past in such a way that it can also serve us in the future.

Cole reiterates Davydov’s view that there has not yet been close cooperation on an international scale between those who take activity theory as their point of departure for analysis and those who use some version of Vygotsky’s ideas with the emphasis on language and semiotics as mediation. He also comments that Russian adherents of these approaches have a different perspective on the differences which are bound up with the historical and ongoing “wrenching political upheavals of the region”.

Wertsch et al distinguish between the terms “cultural-historical”, or “sociohistorical”, as used by Vygotsky and Leont’ev, and the heading “sociocultural” used by many Western researchers applying their approaches today. They feel that the first two terms are appropriate for referring to the heritage we have gained from the Russian psychologists, but that “sociocultural” expresses more closely the ideas as they have been appropriated in current Western research. For one thing, contemporary “sociocultural” scholars do not assume the “universalist” evolutionary claims for a linear progression of the development of the human psyche, which were common to Soviet thinking of the Vygotskian era, though not to Vygotsky himself. Also, a focus on consciousness and the internal workings of the mind was regarded with deep suspicion as “bourgeois idealism” by Soviet officials in power at the time. This was one of the reasons why Vygotsky’s work was banned for so long in his country.

As I have explained, Soviet psychology, grounded in Marxism, has a very different basis to Western psychology. Certainly, too, culture is understood differently in Western and Eastern perspectives. However, a focus on action and on mediation, common to the thinking of Vygotsky, Leont’ev and other Soviet psychologists, is considered to be defining of contemporary sociocultural research. However, Wertsch et al acknowledge that the roles of action and of mediation in human functioning are problematic. For example, the cultural tools actually invoked in an activity, were not necessarily designed for that purpose.

Instead they may have been “dictated” by other sociocultural forces–and their benefits on mental functioning incidental. For example, technology developed for military purposes has had enormous benefits for the development of civilian systems of communication and has facilitated innovation in other fields such as engineering and medicine.

The goal of sociocultural research has been defined as being to:

- Explicate the relationships between human action, on the one hand, and the cultural institutional, and historical situation in which this action occurs, on the other.

Such relationships must include ambiguities, subject to interpretation and debate. This allows for multiple points of view and openness in investigation and analysis. These are essential if theory is to provide any tools for addressing human problems. My approach acknowledges these ambiguities and complexity of perspectives.

AN INTERPRETATION OF VYGOTSKY'S LEGACY

BACKGROUND

Vygotsky, as he would have been the first the admit, was a product of the culture and history of his time. Born in 1896 to a middle class Jewish family in a provincial Russian town and living only until 1934, he was an intellectual, moving comfortably from the realms of art and literature to science and psychology. In Russian culture at the time there were no sharp divisions of labour among science, art, philosophy, literature and even theology. Scholars such as Vygotsky were simultaneously"connoisseurs of these spheres of human activity".

Vygotsky was greatly influenced by philosophers, scientists, psychologists and writers: Spinoza, Hegel, Engels, Köhler, Freud, Piaget, Shakespeare and the poet Mandelshtam, to mention but a few. The teachings of Marx and Engels had a profound effect on his thinking. Unlike many of his colleagues and followers, however, he did not reify Marxism, and ultimately did not accept that Marx' theory provided the explanatory principle of psychology.

This led to his falling into disfavour with the authorities towards the end of his life. His works were banned, some posthumously, and much of his later and most important writing was only published in Russia more than twenty years after his death. Vygotsky saw psychology as experiencing a crisis which was both deep and global. In his view this was a methodological crisis, stemming from the contradiction between the factual material of science and its theoretical premises.

He stated that:

- As long as we lack a generally accepted system incorporating all the available psychological knowledge, any important factual discovery inevitably leads to the creation of a new theory to fit the newly observed facts.

He tried to resolve this crisis by seeking a means of analysis specific to the field of psychology - a systematic way of organising the content, and delineating a framework for psychology.

To Vygotsky:

- 'General psychology' has its own laws, forms and structures.

In his, ultimately unsuccessful, search for the ideal analytic method, Vygotsky became inspired by Marx' analysis of political economics. In the atmosphere of the early 1920's in Russia, there was a struggle between psychology, which used introspective methods and focused on the inner world of the individual, and the study of physiological processes, such as reflexology.

According to Vygotsky:

- Reflexology, which has translated associationism into the language of physiology, sees the intellectual development of the child as a gradual accumulation of conditioned reflexes; and learning is viewed in exactly the same way.

Vygotsky was critical of this reduction of psychological processes to reflexes of various complexities. Reflexology and its opposing physiological theory, reactology, which was formulated by Kornilov and proclaimed the reaction as the basic unit of psychology, threatened to sweep away"subjective, empirical psychology", as it was then in Russia. Marxism was seen by Vygotsky as the saviour of Russian psychology. However, he had his own perspective on Marxism. Vygotsky stressed that scientific truth concerning the mind was not a monopoly of Marx and developed his theories in dialogue with Marxism. This approach differed heretically from ideologically approved perspectives and publications of the time.

Vygotsky's short life span, ill health and prodigious talent meant that, although he espoused many ideas, these did not represent a complete psychological system - they expressed an approach rather than a solution to the problems of psychology. Van der Veer and Valsiner point out that the very lack of conclusion of Vygotsky's theories makes them useful for modern researchers who can develop them in their own settings. Today, there are many theoretical expositions which extend Vygotsky's ideas.

For example, there is considerable research on how cognition develops in social practice, such as Scribner's extensive work; research on situated learning in the form of legitimate peripheral participation; on the role of signs as mediators of cognition as well as"sociocultural" studies.

Indeed, Van der Veer and Valsiner contend that:

- The theoretical debate about the validity and fruitfulness of Vygotskian ideas is only at its very beginning.

In the West, the focus of recent research in education has been on a small selection of Vygotsky's concepts from the plethora of ideas he expressed. This is not surprising in our time of increased specialisation in areas of study. For example, Vygotsky is described by Wertsch et al as a"psychologist, semiotician and pedagogical theorist".

These areas are usually considered quite distinct in contemporary Western disciplines. Also, some of the research work done by Vygotsky and his colleagues and students has been criticised as inappropriate for our political and temporal era.

For example, one claim ascribed to the cultural historical school is that contemporary aboriginal peoples are similar in development to primitive peoples of the Prehistoric era. This criticism refers to studies carried out on so-called primitive peoples in Central Asia by Vygotsky and his close associate, Luria, an important Soviet psychologist. Whether Vygotsky held that view in this simplistic way is, however, considered controversial. One interpretation is that Vygotsky considered the"huge wealth of vocabulary" of these rural people to be an impediment to their developing psychological tools.

Such tools were, however, needed for the development of technology by cultural groups who had less abundant language. More importantly, such criticisms miss the point of Vygotsky's underlying thesis and its great strength - that psychological processes are not racially innate, or formed in the individual's head, in isolation, but develop in the cultural milieu of the individual.

Luria and Yudovich express this idea strongly in their promotion of Vygotsky's approach as follows:

- Only by understanding that the sources of all complex mental processes do not lie in the depths of the soul, but are to be found in complex forms of human social life and in the child's communication with people surrounding him, can we finally outgrow the prejudices which have been rooted for centuries in psychological science.

In contemporary education literature, Vygotsky's contribution concerning the importance of culture to formal or school learning is increasingly being recognised. In what follows, I will interpret some of Vygotsky's ideas, which are relevant to my approach. An important and fundamental aspect of Vygotsky's reasoning is his dialectical thinking. Based on Hegel's philosophy, dialectical thinking involves the resolution of disconnected or mutually discordant points of view–it is a rational method of synthesising conflicting ideas. To Vygotsky, opposing points of view meant that problems were being subsumed. These, in resolution, would lead to new insights.

Van der Veer and Valsiner argue that:

- The over-riding concern evident in Vygotsky's intellectual work is the quest for synthesis.

Hence, though presented separately, Vygotsky's insights are interdependent.

SOME OF VYGOTSKY'S INSIGHTS

Human Evolution as Cultural-Historical Development

Perhaps the most well known contribution of Vygotsky to modern thinking in the West is his emphasis on the "historical development of humans". This "cultural historical theory", in summary, suggests that human higher processes developed in human history and have to be learnt anew by each child by means of social interactions. Vygotsky accepted Darwin's idea that the evolution of man from animals was a continuous process.

In marrying Darwin's evolutionary theory and the Marxist concept that Man has control over his own economic destiny, Vygotsky interpreted the continuity of biological evolution and cultural human development in a dialectic way. That is, he believed that the qualitative differences between humans and animals–the development of higher mental abilities–arose in a complex interaction and antithesis of biologically produced, genetic factors and social interaction.

However, he assigned the major role in the development of particularly human traits to the mastering of culture, particularly the mastering of cultural artefacts such as sign systems and speech. For example, Vygotsky proposed that children's intellectual growth was contingent on their mastering the social instrument of thought, speech–via inner speech.

He called socially developed thought: "verbal thought", which he explained as follows:

- Verbal thought is not an innate, natural form of behaviour but is determined by a historical-cultural process and has specific properties and laws that cannot be found in the natural forms of thought and speech. Once we acknowledge the historical character of verbal thought, we must consider it subject to all the premises of historical materialism, which are valid for any historical phenomenon in human society. It is only to be expected that on this level the development of behaviour will be governed essentially by the general laws of the historical development of human society.

That is, complex processes such as thinking have their bases in the sociohistorical development of human society. To Vygotsky, following Engels' writings, history is the history of cultural artefacts which allows humans to control their own environments. These cultural means, as they are mastered by the developing child, change not only the content of the child's thinking but the very way of his thinking. This means that growing into a culture–appropriating its means–creates in the child a "second nature". To Vygotsky and his colleague Luria, this adopting of cultural modes was not like putting on clothing, because cultural layers, once taken on, cannot be shed. For Vygotsky, who ignored learning at home as informal learning, the means by which this enculturation takes place is education.

His empirical studies focused on children's acquisition of "scientific concepts", that is, abstract or generalised concepts and looked at "the interaction of development and instruction". Vygotsky's view of a rational man, one who is in control of his intellect and behaviour, was that he is an educated man–where education is the process of mastering the means provided by the culture. Vygotsky's definition of culture was selective, emphasising those aspects of human life which enable people to control their environment, such as tools and communication systems–by means of which productive labour is possible. This was not because Vygotsky lacked interest in aspects of human evolution such as wisdom, music, or creativity.

Initially at least, he was inspired to frame his thinking in Marxist terms. Hence the idea of the tools of labour as the means by which people collectively master the physical world and gain economic power, influenced his thinking. The idea of mental tools developed from this notion of man using physical tools to conquer his own destiny. Vygotsky saw signs, for example, words and symbols as analogous to tools in mediating thinking. Van der Veer and Valsiner suggest, too, that Vygotsky would have seen little evidence for progress in areas of social evolution, such as wisdom or creativity. On the other hand, technological artefacts and communication systems–language, writing and symbols–were transforming the ways people functioned.

Hence a focus on these "hard" aspects of cultural development was consistent with Marx' and Engels' characterisation of human qualities as dependent on mastering tools. It was also consistent with Vygotsky's "sincere wish" to liberate and improve the quality of life of all citizens. By having access to education and the mediational means it provided, people could progress in their thinking and so, in Vygotsky's view, free themselves from the yokes of feudalism.

It is possible that a nucleus of Vygotsky's theory of cultural transmission can also be found in his own, absorbed, religious heritage. Vygotsky's religion, Judaism, is rich in symbols and rituals. Further, each symbol, which is adopted by the community, and each ritual in which the community participates, contains the history of the Jewish culture. Also, each shared ritual and cultural symbol or artefact, is itself the seed of continuity–a means of passing on the history and culture from one year to the next, from generation to generation. This is precisely the point of Vygotsky's theory of cultural historical transmission.

Developmental Explanation of Human Mental Processes

Vygotsky's stated aim was to achieve a complete and dynamic analysis of higher psychological systems: their development, their structures and their functions.

He differentiated between the maturation of basic biological functions and the development of higher mental facilities as follows:

- *Within a general process of development, two qualitatively different lines of development, differing in origin, can be distinguished:* The elementary processes, which are of biological origin, on the one hand, and the higher psychological functions, of sociocultural origin, on the other. The history of child behaviour is born from the interweaving of these two lines. The history of the development of the higher psychological functions is impossible without a study of their prehistory, their biological roots, and their organic disposition.

So, in order to explain, rather than merely describe, mental processes, Vygotsky insisted that their origins must be examined: phylogenetic; ontogenetic; microgenetic.

Scribner asserts that to Vygotsky:

- The historical research of behaviour is not an additional or auxiliary aspect of theoretical study but forms the very basis of the latter.

Vygotsky tended to use the terms cultural and historical interchangeably and expressed a need to search for explanations of human behaviour, as opposed to that of animals, in history rather than biology. In his view, higher mental functions are collectively produced and have a sociohistorical, rather than biological, origin. So for phylogeny, the cultural development displaces the biological, that is, the "social" line takes over.

The biologically derived term "ontogeny" is used by Vygotsky as a generic term to refer to all processes of child development, so that in his work the concepts of the cultural and the biological development of the child are fused. According to the Vygotskian perspective, language is the main agency by which society changes biological development to social progress. Vygotsky's understanding of history, both societal history and child development, was not a simple account of universal and linear progress. Rather, he recognised historical processes as changes that move humans in many directions and on many levels. Vygotsky proposed that the relationship between revolution and evolution, though not universally recognised, is understood in scientific thinking.

He expressed this as shown below:

- To the naive mind, revolution and evolution seem incompatible and historic development continues only so long as it follows a straight line. Where upheavals occur, where the historical fabric is ruptured, the naive mind sees only catastrophe, gaps, and discontinuity. History seems to stop dead, until it once again takes the direct, linear path of development.
- Scientific thought on the contrary, sees revolution and evolution as two forms of development that are mutually related and mutually presuppose each other.

Vygotsky's understanding of the relation between history as change and history as unilateral progress is less clearly demonstrated in the work of Vygotsky's students and followers. As outlined, Vygotsky's theories are a powerful lens for viewing the development of the child. His basic claim was that there are two lines of development in the child.

One is the biological growth and maturation of the child. The other, which guides the first, is the mastering of cultural artefacts, particularly speech. The child develops through the internalisation of activities carried out while interacting with others, usually adults. This theme has its roots in the work of Marx as well as taking account of the work of important psychologists of the time, including Piaget and Janet.

Origins of Higher Mental Functions in Social Life

The theme described above: That mental functions emerge from shared social experiences–is formulated by Vygotsky's explanation that a child's development appears first on the social or interpsychological plane then on the personal or intrapsychological plane. That is, Vygotsky proposed that every higher mental function, as it develops in the individual, goes through an external stage in its development because it is initially a social function. In particular, adults' verbal interactions with the child underlie higher functions, such as voluntary attention or concept formation.

According to Vygotsky:

- We could therefore say that it is through others that we develop into ourselves and that this is true not only with regard to the individual but with regard to the history of every function.

Vygotsky described four major stages in the child's developing cognitive processes. These were based on his empirical studies of children carrying out various experimental tasks. In the first stage the child responds to the immediate stimuli around him and his own basic needs.

The beginnings of mental mediation arise in the second stage when the child, instead of solving a problem impulsively:

- Now solves it through an internally established connection between the stimulus and the corresponding auxiliary field.

At this stage, however, the child has not fully mastered the means of mediation. The third stage is reached when the child becomes aware of the role and utility of the mediator in cognitive activity. For example, in one of Vygotsky's experiments, children use pictures to decide which keys should be pressed to fulfil a given task. At this stage children can organise their environment but their regulation of their own behaviour is still dependent on the stimuli present in the external environment. The final stage in the development is characterised by the internalisation of the words or other signs and the relationships among the stimuli, mediator and responses. The operations can now be carried out without external stimuli.

To me, the most important point about Vygotsky's account of this progression to self regulation is his emphasis that this is a social process. Díaz, Neal and Amaya-Williams specify four ways in which this is so. Firstly, the signs and tools that are brought to the setting, and are taken over by the child to control his or her behaviour, are social. They are features of the society or culture at large.

Vygotsky gives many examples of"psychological tools" used in this way, including:

- Language; various systems for counting; mnemonic techniques; algebraic symbol systems; works of art; writing; schemes; diagrams, maps, and mechanical drawings; all sorts of conventional signs.

Secondly, these signs are used by the child to influence others and interact with them - they have an immediate social role. Thirdly, in Vygotsky's view, speech is the most useful agency by which children master their environments. Speech is inherently social - it carries shared meanings with others in a particular setting. Speech not only mediates social interactions, but is also a tool for planning, organising and monitoring behaviour. The child's symbolic activity, especially private, internal speech allows the child independence from the external stimuli and enables mastery over her own responses.

Vygotsky expressed the evolution of"verbal thought" from simple generalisations to the most abstract concepts in this way:

- It is not merely the content of a word that changes, but the way in which reality is generalised and reflected in a word.

So the use of language and other semiotic systems transforms thought in ways that connects the child to her social environment. These transformations are not only in what the child thinks but also in the way she thinks. For example, the strategies people use to solve problems, the categories through which they organise their experiences and the ways people express themselves in language, depend on their cultural history and society. One of the mechanisms by which Vygotsky explained how individuals realise their potential is his construct of the zone of proximal development. This relates to Vygotsky's aim of empowering people through instruction. It also draws on the Marxist notion that individuals can shape their worlds and develop their higher mental faculties through cooperating and interacting with others. Referring mainly to children, Vygotsky defined this zone as the area between what a person can achieve on her own and what she can achieve with the help of a more capable person. This emphasises the child's potential to develop with strategic help. This support is called scaffolding by some researchers, following Bruner emphasising the support given to the learner. It is termed amplification by others which focuses on the child's own potential and current strengths. The implication of the zone of proximal development is that the inexperienced learner will in the near, or proximate, future master a task for which she currently requires some help. Vygotsky chose the word"zone" to underscore the idea that development is a"continuum of behaviours" not a point on a scale. The zone of proximal development has a dynamic nature. As the child masters new and more difficult tasks, the zone of proximal development shifts, and the cycle of assisted learning followed by independent achievement is repeated.

This notion of Vygotsky's is well known and has been extensively researched in the West. It has important implications for mathematics and statistics education where, often, the focus is static - on what students can do unaided at a point in time - the lower boundary of the zone of proximal development. There may be little focus on the continuum of students' skills and abilities or on their potential to progress by means of structured interactions. Traditional assessment in mathematics and statistics reflects the view that a student "knows" or"does not know" a particular concept. Vygotsky's idea of a concept is dynamic. In his term a concept is a maturing process, rather than a mental acquisition

Vygotsky’s perspective recognises the primacy of social contact as a basic human need and also rejects the dualism between the individual and society, insisting on the inseparability of the individual and the social realm in which she acts. The application of these ideas is well documented in the work of contemporary researchers such as Lave and Wenger; Rogoff; Scribner and Walkerdine. Vygotsky and Leont’ev both stress the importance of the social role in human development and learning. However, while Vygotsky tended to emphasise immediate social relationships Leont’ev’s focus was on the broader societal context.

Mediation of Thought

A major goal of Vygotsky was to develop a psychology able to account for the nature of human consciousness. Influenced by his knowledge and interest in art and literature, Vygotsky extended the Marxist concept of physical tools as mediating devices, to mental tools.

To Vygotsky:

- The most essential feature distinguishing the psychological tool from the technical tool is that it directs the mind and behaviour whereas the technical tool, which is also inserted as an intermediate link between human activity and the external object, is directed towards producing one or another set of changes in the object itself.

That is, psychological tools are aimed at mastery or regulation of oneself or others, rather than controlling the physical environment. In a dialectic way, these human transformations lead to the need for further mental tools which again give rise to new ways of conceptualising and acting in the world. Mental tools, for example, scientific theories, are cultural. They are invented and passed on to others through social processes and socially structured practices.

According to Wertsch et al they:

- Provide the link or bridge between the concrete actions carried out by individuals and groups, on the one hand, and cultural, institutional, and historical settings, on the other.

In Vygotsky's writing, the main mediators are words, symbols and other signs.

In this context, a sign has:

- A definite meaning that has evolved in the history of a culture.

A sign can serve as a system of reference, such as "notched sticks and knots", or writing. It can also make possible new, higher levels of thought and analysis, as do, for example, algebraic systems. Vygotsky was particularly interested in the role of the sign, especially the word, in the development of higher mental functions.

To Vygotsky:

All the higher psychic functions are mediated processes, and signs are the basic means used to master and direct them. The mediating sign is incorporated in their structure as an indispensable, indeed the central, part of the total process. In concept formation, that sign is the *word*, which at first plays the role of means in forming a concept and later becomes its symbol. So to Vygotsky, signs enable higher mental processes, such as concept formation, to develop and they also signify meanings. Vygotsky and his colleagues and students, including Luria and Leont'ev, carried out many empirical studies on children and on rural peoples in order to understand how the processes of internalisation of speech and other semiotic systems allowed thinking to develop.

Luria and Yudovich posit that:

- The word, handing on the experience of generations as this is incorporated in language, locks a complex system of connections in the child's cortex

> and becomes a tremendous tool, introducing forms of analysis and synthesis into the child's perception which he would be unable to develop by himself.

That is, the internalisation of word meaning, which is socially developed, provides the structure for the development of cognition. Mental tools, in Vygotsky's thinking, extend the mind. They allow humans to do more than react to the environment as animals must–they enable us to change our environments. This implies that human behaviour is purposeful.

We use tools, and signs, especially language in order to carry out our plans, to cooperate with others and to regulate our own cognitive processes, such as focusing our attention or remembering, in considered, deliberate ways. In Vygotsky's thesis, cognitive functioning is shaped by the sociocultural setting. Mediation by means of cultural tools serves as the mechanism for this shaping.

A PERSPECTIVE ON LEARNING

I review the literature interpreting Vygotsky's approach and Leont'ev's modifications of his ideas. I try to give a range of perspectives on their theories by including the views of contemporary Russian theorists, such as V. V. Davydov and V. P. Zinchenko, and some well known Western scholars, such as J. V. Wertsch and J. Valsiner. To clarify my underlying theoretical position, I differentiate between theoretical approaches which are framed primarily in terms of Vygotsky's concepts and those which take the viewpoints of Leont'ev and his colleagues in the Khar'kov school.

Initially, I explain the background to my literature review and to my application of activity theory. I explain some important ideas of Vygotsky, particularly the cultural historical theory and other aspects of his framework associated with this theory. I develop the tenets of the activity theory approach: its historical emergence, its features and its principles and how these ideas are understood in contemporary research and education.

I illustrate aspects of this theory with examples from my practice. These examples suggest that activity theory has important implications for education research and for understanding students' learning in practice. I comment on unsolved problems of the theory and some contemporary derivatives of the theories of Vygotsky and Leont'ev. I conclude by explaining the relevance of the theory to my investigation–how my approach builds on notions of activity theory and where it departs from these ideas.

BACKGROUND

Activity theory, in the view of Engeström "is one of the best kept secrets of academia". Engeström speculates that this is partly due to language difficulties, in part because of the epistemological roots of the theory in German philosophy and dialectics and also due to the complex and "impressive" theorising behind the activity approach. Yet this theory is unquestionably evolving in Western thought as well as in its birthplace and has shown itself to have considerable implications for practice and research. I develop the theoretical framework introduced. I review the principles of activity theory as they have been explored in the literature. Activity theory has been interpreted differently in different countries and at different historical periods. For instance, the translation of the original Russian scripts into various languages communicates different nuances of the theory. For example, the German word, “Tätigkeit”, evidently conveys a sense of serious purpose, such as work activity or occupation, which is lacking in the English word, activity.

Clearly, too, the perspectives of those who look back at a theory, which emerged in a period of tremendous turmoil and social change, will differ from the perspectives of those who wrote while they were embroiled in the events. Finally “theories in ‘use’ are always theories developing”. I have tried to piece together current interpretations of activity theory and its recent developments from the latest English publications available in Australia. The fact that a number of these scholarly works have been published since 1990 is an indication of the ongoing interest in activity theory and its relevance.

In Western education, the focus of the contemporary activity approach is on two major and related strands. One aspect pertains to the emphasis on goal driven actions as central to a student’s learning. This aspect relates to Leont’ev’s construct of activity. Secondly, based on Vygotsky’s expositions, which are more familiar to Western educationalists, current approaches stress the cultural, historical and social base of these actions as being germane to the learning that takes place.

My aim in this stage is to show how these cornerstones of the contemporary activity approach were developed in the work of Vygotsky and re-interpreted by Leont’ev and other of Vygotsky’s colleagues and students. To begin with, I expand on my interpretation of students’ learning of Statistics, taking aspects of activity theory as my starting points.

USING THE LENS OF ACTIVITY THEORY TO UNDERSTAND STUDENTS’ LEARNING OF STATISTICS

I outlined my approach stemming from my belief that a student’s learning proceeds through her active engagement with the task within a social and cultural setting. This draws on Leont’ev’s notion of activity as “a system in the system of human relations”. A focus on learning as activity which is socially and culturally situated is bound up with exploring what Varela et al term “histories that are lived”. Leont’ev elaborating on Vygotsky’s ideas, distinguished between three kinds of histories that are lived–evolutionary, societal and individual histories.

Firstly, evolutionary history has resulted in humans having bodily structures which enable us to interact with the world. This physical structure outlines the essentials of human existence. Secondly, from a societal perspective, the embodied person is spatially and temporally situated. The self speaks “not only as an individual voice but also as a collective voice", reflecting values and beliefs of communities and cultures which are significant to the individual. Finally, each person develops a personal story, a history of her own experiences which have left their mark. These three histories that are lived, act in tandem in human activity–such as the activity of learning Statistics. The actions a student takes to learn Statistics involve that student's cognitive abilities, which have developed over the history of the human race–her evolutionary history.

This history is inseparable from the development of tools, language and other cultural means of mediation. Secondly, a student's activity is related to her situation in an institutional and societal setting. Thirdly, and interwoven with these other histories, is the personal narrative of the student, both her biological and social development, which contributes to her ways of understanding Statistics and of interpreting the context in which she encounters it.

A key argument of activity theory is that personally meaningful goals have an "energising" function. Varela et al specify two aspects of goal directedness or intention.

These correspond to:

- What the system takes as its possibilities for action to be and to how the resulting situations fulfil or fail to fulfil these possibilities.

In my investigation, this means that actions taken to learn Statistics are directed by how the student construes her goals in context. In turn, how these goals become fulfilled, or fail to become fulfilled, is affected by the conditions surrounding her actions. Goals, with accompanying evaluations, are interpreted within an institutional and broader cultural location. Hence the principle that learning is purposeful, or goal driven, links the two aspects of activity theory that I have delineated–the sociocultural situatedness of activity and the directed actions comprising the activity.

In summary, interpreting a student's learning from the activity perspective means attending to the "system of activities" or practices in which the student is engaged and understanding how the student is oriented or positioned with respect to these particular practices.

Evans and Tsatsaroni describe the positioning of a student engaged in a learning practice as:

- The resultant of the practice(s) in which all subjects in that situation *are positioned*, and of the practice(s) which the particular subject *calls up*.

This implies an operative interplay between participants and setting. This implies an operative interplay between participants and setting. Further, Evans and Tsatsaroni argue that there are critical aspects that make the practices, in which students are engaged, meaningful. These include: the students' goals; the resources available and the constraints confining the students; the social relations, including the exercise of power on which these relations are based and the"discourses which shape the practices themselves".

Their argument is based on suppositions which include both Vygotskian notions, such as the importance of speech and interpersonal relationships, and

Leont'ev's emphasis on the goal directedness of all actions and on the conditions which determine the operations. Hence it is important for me to explain the roots of these ideas, and the others I have outlined, in Vygotsky's and Leont'ev's theories.

FRAMING THE RESEARCH QUESTION

Guided by activity theory and within the methodological framework outlined, I explore the qualitatively different ways students relate to their learning of a statistics course at university. Statistics, with a capital"S" refers specifically to the statistics component of the Psychology II course. Hence this usage immediately specifies the setting of the activity as university learning of a particular topic at a defined time and place.

The main research question which I investigate is:

- What are students' orientations to learning Statistics ?

I am interpreting the word "orientations" as signifying students' positioning of themselves with respect to the learning task. This question is important because, from the theoretical perspective I have outlined, students' orientations to learning Statistics are integral to the ways students engage with the learning task and hence relate to the quality of their emerging knowledge.

To investigate the research question I explore three interdependent aspects. Firstly, I investigate different facets of students' orientations to learning Statistics, including affective elements, students' conceptions of Statistics and their approaches to learning the subject.

Secondly, I look at the connections between students' orientations to learning Statistics and the outcomes of learning it, mediated by their activities. Thirdly, investigating this question intrinsically includes exploring the relationships between the students' perceptions and actions and the contexts surrounding their learning. For example, if a student's conception of Statistics is "algorithms", this is not a characteristic of the student or of the subject Statistics or of the educational institution.

It is a relationship between the student's way of experiencing Statistics and the contexts surrounding these experiences. From the activity perspective, how a student orients herself in the learning arena her actions and her evaluations of the learning task, develop together and coherently and, with their accompanying outcomes, are organised within and contribute to the wider sociocultural setting. A student's orientation to learning Statistics could be regarded as a particular mind set individually generated in response to her experience of a certain context.

My view of orientation is in contrast to this view of it as an individual and internal construction. Consistent with an activity theory framework, I regard a student's orientation to learning Statistics as inseparable from her actions to learn Statistics and as part of a wider and dynamic societal "system of activities". Statistical thinking cannot occur in isolation, in the head.

In learning Statistics, an individual is participating in a cultural practice even when she appears to be acting in isolation, such as doing examples or reading a text book. Further, the statistics that the student is studying has itself developed historically through interactions which are culturally founded. The methodology for statistical inference, for example, has been cultivated and improved by successive generations and continues to undergo major transitions as a result of current technology.

My interpretation of the activity of learning Statistics, drawing on Leont'ev's sense of the construct, refers to the actions a student takes to grapple with it–to engage, purposefully, with the problem of learning it. A student's engagement with the task of learning Statistics is expressed by both cognitive actions and practical actions. These actions are directed, not random, even if the student does not articulate her goals. By "doing" Statistics a student "makes" it "artistic, practical, creative or routine".

CONCLUSION

The literature on statistics education shows a great concern for improving teaching. My way of approaching this problem is to try and increase understanding of how students learn Statistics. I have introduced Leont'ev's theory of activity and my application of it to this understanding. In summary, I propose that students learn through their actions in context. That is, my interpretation acknowledges that learning Statistics is a human activity that occurs in a social, cultural and historical context. My perspective stresses the changing and fluid course of a student's statistical thinking–actions realise the emerging statistical knowledge. Moreover, the meaning of Statistics to the individual does not produce her statistical thinking–it mediates her thinking in the same way as tools mediate rather than produce changes in the environment.

4

Experiments in Research

Typically, our exposure to formal experiments begins in the context of natural or life sciences and is continued in the news media in the context of medical research. The word “experiment” is a commonly used term. As students we are exposed early to the notion of experiments and most of us have conducted experiments. By the early years of high school we conduct experiments in chemistry classes that measure the effect on one chemical caused by adding another chemical.

The word experiment is heard regularly in the general media in connection with the study of some action or treatment. It is often in the form of news accounts reporting results of medical experiments. An example is a news report of an experiment conducted by the UCLA Medical School to test the effect on reducing heart attack caused by taking an aspirin a day. Another example is an experimental test by the Upjohn Pharmaceutical Company regarding the effect caused on growing hair on a man’s bald head by applying Minoxidil (trade name Rogaine).

Experiments are conducted frequently in marketing research, as well as in research in medical fields and the social sciences. Regardless of discipline or application, an experiment is a test of a hypothesis that a causal relationship exists between two or more variables.

This test focuses on an effect or change that one variable is believed to cause in another variable. The word experiment is used most frequently in the English language as a noun referring to a study or an investigation of a problem or question. The word experiment is also used as a verb referring to the action of conducting a study or investigation of a problem or question. People who are studied or investigated in an experiment and from whom information is collected are called subjects, in contrast to people in a survey, who are called respondents.

A subject is a participant in an experiment, from whom information is collected. Subjects represent some larger group of people to whom information from the subjects is generalized. This larger group of people is likely a target market in which the user of marketing research is interested.

While surveys are the most frequently used method of investigation in marketing research, experiments are also used frequently. Experiments are used in investigations of all the controllable marketing mix variables (*e.g.*, product, price, promotion, and distribution), consumer behaviour, and even in regard to marketing research itself.

These experiments often incorporate uncontrollable variables of marketing such as the economy and competition, and do so in different applications of marketing such as international marketing, business-to-business marketing, and services marketing.

TEST MARKETING

Experiments in marketing research are especially prominent in test marketing in order to decide if and how an organization should market a new product. Test marketing is a study that obtains response information through use of an experiment in a realistic marketing setting. While numerous measures of consumer acceptance are conducted (such as product taste, product appearance, package size, price, *etc.*) before a product is introduced into the actual marketplace, test marketing is the final measure. Experiments for test marketing have proved valuable. One report stated that about 75% of all new products that were test marketed were a success in the actual marketplace, compared to a success rate of only about 20% for new products not test marketed.

Despite the past popularity of using experiments in test marketing, test marketing is the one area in which continued popularity is debated. One survey of 1,200 firms reported that 35% used test marketing in 1975, and 12 years later, in 1987, only 30% used test marketing. The reason is related to the time, money, and less-than-perfect information associated with such applications of experiments. However, these considerations pertain more directly to test marketing than to the topic of using experiments in marketing research.

FUTURE

The future of experiments in marketing research is bright. An article published in 1990 reported that between 1975 and 1987, the number of firms using “formal experimental designs” increased from 16% to 42%. During this time, firms using “informal experimental designs” increased from 35% to 57%.

There are strong reasons for these increasing statistics. To begin, marketing practitioners constantly encounter questions about cause-and-effect relationships regarding the controllable variables that comprise the marketing mix. These questions often involve specific marketing efforts, concerns about attitudes, and desire for sales. These issues align well with the major components of experimentation, such as hypotheses, independent variables, dependent variables, and conclusions about causality that are not possible with research methods such as observation and survey.

Experiments in marketing research are conducted in regard to every aspect of the marketing mix. Experiments are utilized to investigate product concept, product composition, product appearance, package sizes, Colours, prices, distribution outlets, promotion alternatives, advertising appeals, advertising copy, overall effectiveness of different commercials, proper store shelf placement, and on and on. They also are utilized to investigate marketing functions such as professional selling techniques and sales management practices, and even to determine the best number of business-to business (industrial) sales calls to make.

In fact, whenever company management has a question about specific marketing management alternatives, such as one package size versus another, an experiment is likely applicable for investigating the answer. Therefore, it is reasonable to expect their use will increase in all areas, with the possible exception of test marketing. The reason is that company managements are inundated with cause and-effect questions related to marketing mix variables, when, at the same time, no other research methodology provides as conclusive information about such causal relationships.

Experiments offer two key features not offered by other research methodologies. These are (1) the ability to actually investigate causal relationships and (2) to exert control over major components of the research project. These two features are interdependent such that the ability to investigate causal relationships is due to the extent it is possible to exert research control.

CAUSALITY

Causal relationships are the heart of experiments. Regardless of the distinction between people who conduct marketing research and those who use its results, mention of an experiment communicates to everyone the notion of an investigation of the relationship between a particular cause and a particular effect. This notion is in contrast to how potential relationships are investigated through observation and survey research methodologies.

Causality is a relationship in which a change in one variable causes a change or effect in another variable. The first variable is referred to as the independent variable and it causes an effect on the second variable, referred to as the dependent variable. Causality is a topic of special importance in marketing research because marketing management is continually interested in causal relationships. But conducting a marketing research project to investigate the effect on sales (“Y”) that is caused by a change in a product or marketing mix attribute (“X”) is not nearly as straightforward as may appear initially.

The cause-and-effect relationship of interest to researchers is commonly expressed by the phrase “X causes Y.” Cause-and-effect relationship (or causal relationship) is synonymous with causality and, in reality, the phrases or terms are used interchangeably. Likewise, "X causes Y" is a phrase synonymous with causality and, in reality, the two phrases or terms are used interchangeably.

Research theory shows that causality is a complex topic. Understanding the complexity at the level of research theory is not necessary at the current level of marketing research discussed in this book. However, it is important to be knowledgeable about two related fundamentals: (1) the basic expression of causality and (2) the conditions required for causality.

BASIC EXPRESSION

The basic expression for causality is "X causes Y." Certainty of causation implied at this level is referred to as deterministic causation-a relationship in which "X" always causes "Y." Among the many misuses of the term "causality," implying deterministic causation is the most blatant.

Consider the common misuse of causality, and even deterministic causation, by the news media. The news media routinely (and erroneously) report causality, and even suggest deterministic causation, based merely on the fact that two events occurred during some time period. A specific example is the media's reports that "today's positive (or negative) effects on the stock market were caused by" some individual event in the country or world. Sadly, even financial experts voice similar proclamations of cause and-effect relationships related to the stock market.

Deterministic causation rarely exists. At best, when causation does exist, it is most reasonably to infer probable causation — a relationship in which "X" is likely to cause "Y." When causal relationships are spoken about in marketing (as well as finance, business, social sciences, biological sciences, and life in general), we are most usually referring to likely or probable causation and not deterministic causation. The difference between probable and deterministic causation is substantial. Therefore, at least three research theory considerations are important to be aware of in regard to the causality expression of "X causes Y":

- Rarely does only one "X" cause "Y." Instead, it is likely that many variables determine an outcome.
- The presence of "X" might increase the likelihood of "Y," but increasing the likelihood is not the same as making its occurrence definite.
- "X" can never be proven with absolute certainty to cause "Y." Based on available evidence it can only be inferred with reasonable certainty.

These three considerations pertain well to marketing and marketing research. First, rarely does only one variable such as package size ("X") cause substantial differences in the level of sales for the product ("Y"). Sales are determined most always by a combination of variables such as the package size, the product itself, its package design, price, promotions, distribution, and so on. This multiplicity of marketing factors equates well with the notion of multiple variables represented by "X1", "X2", and "X3", causing an outcome such as sales represented by "Y."

Second, because a product is marketed in, let's say, the size most desired, which equates with the presence of "X," it might increase the likelihood it will be purchased ("Y"). However, availability of this desired package size does not make its purchase definite, since the marketplace may purchase a competitor's product or may already possess a sufficient quantity of the product. Third, even though sales of a product may have increased ("Y"), as the product's package size was changed ("X"), marketing executives can never be absolutely certain this change in sales ("Y") was caused by the new package size ("X").

Based on the available evidence, the most that marketing executives can do is infer a conclusion with reasonable certainty, since in the marketing discipline there are many factors that influence product sales. For example, at the same time the product was being sold in a new package size, the competitor's product may have been withdrawn from the market due to a defect, perhaps a special promotion campaign was begun by the competitor, possibly the price was lowered, or maybe the economy changed, and so forth.

REQUIRED CONDITIONS

Research theory shows causality cannot be proven or concluded with absolute certainty. It does show, however, that causality can be inferred and subsequent conclusions made with reasonable certainty. According to research theory, evidence of association, appropriate timing, and elimination of alternative explanations are all required to make even reasonable conclusions about causality.

EVIDENCE OF ASSOCIATION

For legitimate conclusions that causality exists, two entities or events must be associated in a manner referred to as concomitant variation. Concomitant variation is the extent to which a cause ("X") and an effect ("Y") occur together or vary together.

While concomitant variation is a required condition to infer causality, it alone is not a sufficient condition. For example, the fact that a change in the container size of Heinz ketchup occurred at the same time as the product's sales increased is concomitant variation. They both occurred or varied together. However, other events in the marketplace could have occurred at the same time as the package size change, and these other variables could have caused or contributed to the change in sales.

APPROPRIATE TIMING

Appropriate timing is required to make legitimate conclusions that causality exists. Specifically, a cause ("X") must occur before or simultaneously with the occurrence of an effect ("Y"). The required order is that a cause cannot occur after an effect.

Consider the situation in which sales increase during the month of June for Heinz ketchup, and afterwards, in September, a new container size of the product is distributed in the marketplace. Obviously, it is not logical to conclude that this increase in product sales was caused by the change in package size.

Sometimes the timing of events is less easily identified. In these situations care must be exercised to determine the order in which the events of interest occur.

For example, maybe customers became aware that a new, less desirable size of container would replace the current desirable size container. In response, the purchasers rushed out to purchase the current size, with the effect of increased sales actually being caused by current anticipation of a future event.

LACK OF ALTERNATIVE EXPLANATIONS

Alternative explanations for what may cause an effect ("Y") must be eliminated if legitimate conclusions are to be made about the existence of causality. If more than one explanation can explain why an event occurred, then a conclusion of causality cannot be made about a specific cause.

Consider that during the month of March a new container size of Heinz ketchup is marketed. Also in March a new advertising campaign is begun on television and a new point-of-purchase display is set up in the grocery stores. Then, in April, the Heinz company experiences a 10% increase in its ketchup sales.

It cannot be concluded that the new package size caused the increase in sales (*i.e.*, the effect) because the advertising campaign and the point-of-purchase display also were new during that time and, therefore, also offer explanations for the increase in sales. In order for a legitimate conclusion to be made about the causality between the package size change and the product sales, it is necessary that all other factors but one remain unchanged during the study of the relationship between these two variables of interest. Being able to eliminate alternative explanations in relationships believed to be causal is of extreme importance to marketing research in particular and business in general. For an example of the relevance in the advertising and broadcasting industry, which demonstrate how this importance translates into ongoing issues with immediate financial consequences.

This advertising example and the above Heinz ketchup example are problems of trying to eliminate alternative explanations after the fact. These two examples illustrate the value of being able to eliminate alternative explanations before the fact. Experiments make it possible to accomplish this task because of a major feature known as "research control."

RESEARCH CONTROL

A hallmark feature of experiments is the ability to make conclusions about causality with reasonable certainty. The aspect of experiments that makes these conclusions possible is the researcher's ability to exert greater research control than is possible with other research alternatives. Research control is the ability to manipulate research procedures in a manner that eliminates alternative explanations for the findings. These manipulations involve the variables believed to cause the effect in which company management is interested.

Company management at Procter & Gamble may believe that the variable of package Colour has a significant effect on sales of toothpaste. It may wonder if changing the current, primarily white container of Crest toothpaste to a subtle green Colour would cause more product sales.

In this example, the consequent research problem, no doubt defined earlier from the product definition stage of the marketing research process, is to investigate if a green container produces a higher level of sales. Research control is an important feature of experiments, regardless of whether the results indicate support or rejection of the belief. Therefore, if the results in this example indicate that a green toothpaste container does increase sales, the researcher could infer with reasonable certainty that a respective conclusion is accurate.

ISSUES IN EXPERIMENTAL DESIGN

There are several additional issues in experimental design that are useful to consider when planning or evaluating experimental research. The experimental design paradigm that has been discussed in this chapter and illustrated in the El- Bassel et al. article is that of the classical group experiment, although one taking place in the field rather than in a laboratory setting. Important variations of this design exist and there are issues in sampling, data collection, and data analysis that are of particular importance in experimental designs.

Variations in Designs

As the discussion of experimental designs has illustrated, an experimental design itself is complex and involves many elements. Key design elements include the use of two groups for comparison, random assignment of participants to groups, repeated measurement and the introduction of an intervention or manipulation. Any one of these design elements may be altered, although changing them may affect the conclusions that can be drawn if confounds are introduced. So far, experiments have been discussed primarily in terms of the comparison made between the experimental and the control groups.

The design can also involve comparisons within groups, however, that is, in each group from the time before the manipulation or intervention was made to the time after. Campbell and Stanley point out that the posttest-only design, as long as it includes random assignment to treatment and control groups, may be just as strong in internal validity as the pre- and posttest design. A posttest-only design involves only between group comparisons; the pre- and posttest design involves both between and within group comparisons.

Other variations, however, have greater effects on the internal validity of the design. For example, a study may examine a change before and after some event that was not in the control of an investigator. Measurements may or may not be repeated. Group membership may or may not be randomly assigned. No group for comparison may be included and only a within-group comparison over time may be made. Since each part of the experimental design is there specifically to eliminate confounds, any of these changes in fact introduces one or more confounds into the design.

For that reason, such designs are termed quasi-experiments because they do not fully adhere in all respects to the logic of a true experiment. Quasi-experimental designs are quite common, especially in applied fields such as education, social work, and human services research. While they confer some disadvantages on the clarity of the conclusions that may result, they can be quite useful, especially if viewed as a form of relational research. When a true experiment cannot be used because the variable of interest cannot be manipulated by the researcher, an analogue experiment may be conducted.

That is, the variable of interest may be manipulated by analogy rather than in real life. For example, the helping professions have been quite concerned about the effects of racism on practice, especially on the assessment of clients. Since the race of a client cannot be manipulated for study, case materials, such as a case history, written vignette, or videotape, can be developed and the race of the client manipulated to see if the social workers' responses to the case are the same or different depending on the race of the client described. The Paviour study of social work students' assessment of cases is a good example of just such a study. These designs are true experiments, but they present a problem of interpretation in that there is no way to know whether or not responses to case materials or a written vignette differ from how the same respondents would react when confronted with real clients.

However, analogue experiments can be useful tools for addressing questions when manipulation is not possible in the real world. Sometimes in an experiment more than one intervention may be made and compared, or the intervention may be varied, thus involving more than two groups in the experiment. Similarly, an intervention may be made in groups that differ systematically on some other variable or variables of interest. What results in either of these cases is termed a factorial design, that is, one in which more than one independent variable, or factor, is compared.

One question that such designs must address is whether each factor by itself seems to have an effect on the outcome variable, termed a direct effect, or whether some combination of the factors may have a unique effect together. These combined effects are termed interaction effects. Sophisticated techniques of analysis are available that can separate direct from interaction effects, and these should be utilized whenever factorial designs are employed.

One variation in design that has already been mentioned is the single-subject experiment. The fundamental differences between single-subject designs and the type of experiments thus far considered is the context of the comparison. In the type of experiments discussed so far, termed group experiments, independent variables or interventions are examined by comparing what is observed in two groups of participants deliberately treated differently. In singlesubject experiments, independent variable effects are tested by comparing what happens in an individual participant before and after the intervention is made. Single-subject experimental designs originated in biology and medicine where the primary interest is in treating the individual case. For this reason, many helping professionals have advocated single-subject experimental methodologies as especially suited to the clinical situation, where treatment is directed towards individuals or individual cases, not towards groups.

Sampling

The hallmark of an experimental design is manipulation: some change is introduced affecting one or more groups of people. Experimental research thus demands quite intensive involvement with and study of participants as well as a high degree of access to and control in the observational situation. For these reasons, samples in experimental studies are often relatively small and/or carefully chosen. Experimental studies thus do not depend, as descriptive and most relational ones do, on large, representative samples. Rather, the ability to generalize from the results of experimental studies depends instead on the strength of the theoretical explanation of the observations made and on replication. The key issue in sampling in experimental research, then, is the way in which assignment of study participants is made to experimental and control groups. Random assignment to experimental and control groups is essential to the logic of the comparison that an experiment is designed to make.

If more than one factor is being manipulated, of course, the randomization procedure must include all the groups. Random assignment can be done in a variety of ways, from drawing numbers to using a random number table or computerized random number generator. What is essential is that it take place in order for a study to be considered a true experiment.

Data Collection

In general, data collection issues in experiments are no different than in any other form of fixed method research. There are only two additional points to be made. The first is that if the experimental design involves a pre- and post comparison, data collection procedures must be used that work well with repeated use. Experiments are no different in this respect, of course, from any other form of longitudinal research. Some experiments are designed to be "blind" experiments; that is, people involved with study participants during the course of the study are not supposed to know whether a given individual is in the experimental or the control group.

Data collection procedures and any data revealing who is in which group must be carefully safeguarded in such circumstances. For example, in Applegate's study of the effects of client self-monitoring on treatment outcome, some therapists learned from the spontaneous remarks of their clients that they were participating in the self-monitoring process.

While the researcher elected to gather and use these unexpected, spontaneously generated data as part of the findings, there is a possibility that a therapist's knowledge that a client had participated in the self monitoring affected the Global Assessment Scale ratings given after eight interviews, which was one of the study's outcome measures. However, the presence of other client-based outcome measures seemed to validate the overall finding of no significant difference in therapist ratings of treatment outcome.

Data Analysis

However carefully an experiment is designed, its observed results rarely reflect the hypothesized outcome perfectly. In most cases, the results of an experiment or a study of treatment in social work and the human services are not expected to be deterministic or exactly as predicted in every case. In reality, the are expected to be probabilistic, that is, likely on average. El-Bassel et al., for example, hypothesized that those clients who participated in the skills-building group would do better one month after release from prison than those who did not. However they did not expect that every client in the experimental group would be rated more positively on every outcome measure than every client in the control group. Logically, given the complexity of human psychological life, it is unreasonable to assume that every person will react in the same way to any event. In addition, the methods available for assessing effects, such as, for example, self-reports of high-risk behaviour, are not so precise and accurate that even if perfect improvement were obtained the questions used to measure it would necessarily capture the change.

The question then is: how likely is it that the empirical results obtained do or do not support the study hypothesis? Often, then, hypotheses of difference are examined at the data analysis stage using inferential statistics. Inferential statistics require that the original hypothesis of difference be divided into two sub-hypotheses called the research hypothesis and the null hypothesis. The alternative hypothesis of difference and the null hypothesis are arranged so as to be mutually exclusive, which means that only one of them can be true.

It is the null hypothesis that is then tested statistically, with the hope of rejecting the null hypothesis and thereby providing support for the alternative hypothesis of difference by implication. Inferential statistics, which are based on assessing probabilities, are the tools used to examine the probabilistic nature of experimental results. The statistical analysis of data from experiments uses a set of inferential techniques for examining hypotheses of difference termed the analysis of variance. When factorial designs are used, it is necessary to employ the multivariate versions of these techniques in the analysis.

Finally, when pre- and post-comparisons are included in an experimental design, the specific statistical tests to be used are different from those used when the groups compared are independent of each other. Many experimental designs employ both within group and between-group comparisons. The statistical procedures used in such a circumstance are termed mixed models because they incorporate the comparison of both independent and correlated groups.

ETHICAL ISSUES IN IMPLEMENTING CONTROL GROUP DESIGNS

Given the power that a well-designed experiment has to demonstrate the effects of manipulated factors on specific outcomes, including interventions or treatments, it is somewhat surprising that experiments are not more often undertaken in social work and the human services. There are no doubt many explanations for this fact. Experiments are very expensive and demanding to conduct, and they are intrusive in method. They also require, of course, that both the intervention being studied and the effects of it that are expected must be specific and specifiable, which requires a level of theory development that is often difficult to attain.

Perhaps the greatest barrier encountered to the conduct of experiments to study practice, however, has to do with an essential element of their design: the control group. Imagine approaching a group of practitioners and asking them to participate in a study of a promising intervention that may help the clients they are already serving. Imagine their reaction when told that as part of the study a control group of potential clients that will not participate in the treatment must be recruited and studied as well. To many practitioners this suggestion will seem an unethical one. How is it possible to defend withholding a potentially useful treatment from someone who wants and might benefit from it? Debate about this kind of study procedure is now, in fact, quite widespread in medicine, especially when trials of experimental treatments for serious and life-threatening conditions like HIV disease are being discussed.

The defence of the use of the control group or "no treatment" condition in an experiment is basically simple: Experiments are only useful when there is a genuine question about the usefulness, or even the potential harmfulness, of a treatment or intervention. If a programme or service is known to be useful, there is not an ethical justification for withholding it. Research, however, is conducted at the boundaries of knowledge, when there is a genuine question about whether an intervention will or will not be effective under certain conditions or for whom. For example, if the El-Bassel et al. study had found that the experimental treatment had strong positive effects on all the outcome measures assessed, it would naturally be hard to justify a future study in which some clients were not permitted to participate while others were. Even so, both theory and preliminary research findings like theirs suggest that solely educational approaches to HIV prevention may not be effective, at least in some high-risk groups.

There are some specific adaptations that are sometimes made to experimental designs in real-world practice settings, however, that can help to reduce potential ethical problems with them. Sometimes an agency or programme has more clients available and seeking its services than it can accommodate at any one time. In this circumstance, some way of selecting clients to receive the service must be found. Random selection of clients into treatment and control groups from a waiting list can work quite well. Every client on the waiting list has an equal chance of receiving the experimental treatment, and those who are not getting it initially could not have been accommodated anyway.

In fact, not using random selection in such a circumstance will often introduce confounds into a study. It is not difficult to imagine examples. Suppose limited funds are available for a study of the effects of a school breakfast programme on children's school performance; would a teacher or social worker selecting participants for the study not be tempted to be sure the hungriest children were included in the treatment group? However, by comparing an experimental group composed of the hungriest children to a control group composed, by default, of the better fed, the groups would not be comparable to begin with.

In fact, in such a circumstance it may become less likely that the intervention will show an improvement in performance because the experimental group will be full of those at greatest risk. Once a study period is over, however, participants in a control group can be made eligible to receive services, only at a later time. In fact, experiments can be planned to provide treatment or services to successive waves of eligible clients. This kind of successive treatment design is termed a staggered baseline design in which those initially excluded are later included and studied as a successive group, and it can be used whether or not a waiting list exists at the beginning of the first study period or not.

Finally, experimental studies often opt for an alternative treatment comparison group rather than a control group. In such designs, a new and innovative intervention is added to or substituted for a standard or traditional one rather than consigning a group of study participants to no treatment or services at all. In fact, this is the approach used in the El-Bassel study. It would have been considered unethical not to offer the standard educational intervention about HIV disease in order to achieve a "no treatment" condition even though there were doubts about its effectiveness.

Whatever the design used, however, once an experimental study has accumulated data that suggest an intervention is truly helpful or if some new effective treatment for the problem under study were to become available from somewhere else, it is always necessary to revisit the question of the use of a control group. It is only ethical to ask people to participate in an experiment with a control group, or even a comparison group, if there is real doubt about the effectiveness or potential harmfulness of the intervention under study. Even medical studies are often stopped in their tracks when the accumulating evidence suggests that the treatment being studied is effective enough that withholding it from some can no longer be considered right.

THE CLASSICAL VIEW OF EXPERIMENTS

Answers to questions about what will happen next have traditionally been interpreted as indicating whether one event causes another on the basis of a logic developed by the philosophers David Hume, John Stuart Mill and Karl Popper. Their logic, summarized below, provides the rules for when it is considered scientifically acceptable to conclude that the manipulation of variable

A "causes" a change in variable B. Because this logic has traditionally been so influential not only in defining experiments but also in defining research itself, it is worth reviewing it in some depth.

The logic of experimental designs is rooted in philosophy, specifically epistemology. The philosopher Hume, for example, considered the conditions that result in the commonsense belief that one thing causes another and identified three. The first is contiguity between events: Using symbols, if A is presumed to cause B, then it must be true that A and B touch one another in time. Hume's first condition clarifies an important distinction between science and philosophy: Science is concerned only with the immediate, not ultimate, causes of things. Hume's second condition also concerns the temporal relationship between presumed causes and effects, and has to do with order in time: If A is presumed to cause B, then A must precede B in time. This second condition at first seems intuitively obvious: Of course event B cannot be caused by event A if event A occurs after it, since that would mean B was caused by something that did not yet exist when A occurred.

Hume's third condition is the most demanding, and the most important: "constant conjunction." Constant conjunction between a presumed cause and a presumed effect means that the presumed cause must be present whenever the presumed effect occurs. By this third condition, if one is to be able to argue that A causes B, then every time B occurs, A must be observed to immediately precede it. Hume's condition of constant conjunction, while pointing the way towards some important logical considerations, was somewhat problematic. Among other difficulties, it was unable to handle events that were multiply caused, such as sneezes, for example, which can be preceded either by inhaling pepper, exposure to pollen or incubation of a cold virus.

It remained for the philosopher John Stuart Mill to elaborate Hume's logic. Like Hume, Mill argued that three conditions must be met in order to justify concluding that one event caused another. Mill's first condition was identical to one of Hume's: The presumed cause must precede the presumed effect in time. Mill's second condition was an expansion of Hume's principle of immediate contiguity. While Hume had argued that a cause and effect must touch each other in time, Mill's second condition was less restrictive: Mill argued that there must be a demonstrably regular temporal relationship including, but not limited to, immediate temporal contiguity between a presumed cause and a presumed effect. Mill's third condition was the most complex and the most important.

It can be stated simply: arguing that A caused B requires ruling out other potential explanations for why B occurred. Mill's third condition, while simple to state, is hard to fulfill: In practice it is generally very difficult to arrange observational conditions that rule out other reasons for an observed change. In addition, in the end an argument that is causal depends on theoretical explanation for what has been observed. Mill articulated three methods for fulfilling this third condition. The first, called the Method of Agreement, requires that the presumed effect be observed to be present whenever the presumed cause is made to occur. Returning to symbols, if A is the presumed cause and B the presumed effect, the Method of Agreement requires that the presumed effect be observed to be present every time the presumed cause is made to occur. Returning to symbols, if A is the presumed cause and B the presumed effect, the Method of Agreement requires that "whenever A, then B" must be observed.

The second method is called the Method of Difference. This method requires that whenever the presumed cause does not occur, the presumed effect will be observed to not happen. Using symbols, if A is again the presumed cause and B the presumed effect, the Method of Difference requires that "whenever not A, then not B" must be observed.

Mill's third method, the Method of Concomitant Variation, ties the first two methods together. According to the Method of Concomitant Variation, when the first and second methods are observed simultaneously, and under otherwise equivalent conditions, potential explanations other than the presumed cause for why the presumed effect occurred can be ruled out. Again using symbols, where A is the presumed cause and B the presumed effect, Mill's argument implies that observing "when A, then B" and "when not A, then not B" simultaneously, under otherwise identical conditions, offers strong evidence that A was the cause of B. Hume also argued strongly that a logical case for a cause and effect relationship can only be made for events in the past. Even when a strong argument can be made that an event caused another event in the past, there is no empirical reason to argue that the same causal connection will hold in the future. While it is common practice to believe that causal connections regularly observed in the past will continue to hold true in the future, there is no logical reason for doing so, and Hume and Mill were careful to make this point clear.

This logical qualification had profound implications for the development of a science for the helping professions. Presumably, the primary aim of a science of the helping professions is to identify causal connections that will enable helpers to produce change.

But change must take place in the future; if identified causal connections are logically relevant only to the past, they are of no use to change agents, or practitioners. Fortunately the philosopher Popper contributed thinking that permits the development of a science of helping. Popper conceded that because Hume's position that observed past cause cannot logically be taken as proof of future cause is true, it is not enough to expect science only to accumulate collections of causal observations in the past.

Popper argued that science must develop an articulated theory that includes certain causal connections reasonably expected to hold true in the future. The role of research, then, is to determine whether specific theoretically derived explanations of causal connections can be observed to be true in reality, that is, empirically. However, it is not the repeated observation that the sun "comes up" each morning that gives us confidence that it will "rise" again; it is the explanation for its regular appearance that is convincing.

In Popper's view, theory is given support to the extent that its implied causal connections are observed to hold true, assuming that those same connections are not more simply, or parsimoniously, implied by a different theory. However, the overarching theory cannot be proven, even when its implied causal connections are observed, since those same connections could potentially be understood or explained by a better but as yet unknown theory. Fallibilistic realism also holds that theory is a critical organizer of understanding and central to the scientific enterprise even though the notion of "proof" or certain knowledge is rejected.

According to Popper, the role of experimental research, then, is twofold. Its first role is to discredit false theories, which it does when experiments fail to produce results clearly implied by the theory under test. Popper termed this role for experimental research "falsification," and although disconfirming data are not always or unambiguously regarded as falsifying the theory from which they derive, the notion that carefully collected unexpected or anomalous results will call theories into question is generally accepted. The second role of experimental research is to provide support for overarching theories, which it does by implication: If data collected in the context of an experiment do not falsify a theory, by implication they provide support for it.

The logic of experiments, then, following Hume, Mill, and Popper, may be seen in traditional terms as following this course:

- The researcher begins by stating a theory. From this theory will be deduced one or more specific statements about an implied causal connection between two phenomena: the hypothesis.
- The researcher will presume this implied causal connection is true and will arrange conditions to demonstrate this presumption.
- The demonstration will involve arranging two sets of circumstances identical except for one feature: the presence or absence of the presumed causal phenomenon.
- After arranging these circumstances, the researcher will observe what happens under them. If the presumed effect occurs when the presumed cause is present and the presumed effect does not occur when the presumed cause is not present, the researcher will conclude that the original theoretical statement has received support. If the presumed effect does not occur when the presumed cause is present, or if the presumed effect occurs when the presumed cause is not present, the researcher will conclude that the original theoretical statement may be false.

These four steps form the underlying logic of experimental design and provide the ground rules for distinguishing between credible and flawed experimental studies. Conducting an experiment, then, involves being able to make two arrangements. First, the experimenter must be able to introduce or withhold the causal phenomenon, termed the independent variable, at will. Only when this is the case is the independent variable truly free to vary, that is, truly independent, so that its presence or absence, but more importantly what is seen to occur contingent on its presence or absence, cannot be held to depend on or result from anything else, such as another, hidden causal variable. Second, the experimenter must be able to arrange, or gain access to, at least two sets of circumstances that in the ideal case are identical or, more realistically, at least do not differ from one another in any important respect.

In this case, what is important is what happens with the phenomenon that is expected to be changed, that is with the dependent variable. It turns out that arranging for the absence or presence of the independent variable, while not always possible, is the easier of the two to arrange. Creating or gaining access to identical or equivalent sets of circumstances, which is essential for making uncontaminated comparisons between "what happens to the dependant variable when, versus what happens when not?" is often quite difficult.

Early on, it was believed that since the best way to compare identical circumstances would be to compare the same ones, the purest comparisons would involve seeing what happened under a given set of circumstances before versus after the introduction of an independent variable event. As will become clearer in a later section of this chapter, however, it cannot be argued that circumstances before and after the introduction of an independent variable "stay the same" except for the introduction of the independent variable. Just because the experimenter has not done anything else to make them different does not mean that they aren't different. And in fact, circumstances before and after the occurrence of an independent variable will always also be different in at least one respect: the after circumstances will always occur at a later point in time. This time difference turns out not to be unimportant. For this reason, it is considered much better to make comparisons across sets of circumstances occurring at the same point in time.

The research exemplar in this chapter includes just such a comparison. Before examining this illustration, it must be noted again here, as it was in the context of discussing other types of research designs, that there is no such thing as a perfect experiment. All experiments involve making compromises between ideal design principles and actual conditions. No matter: An experiment need not be perfect to be valuable. It needs only to be carefully designed, carefully done, and honestly reported.

GENERALIZING FROM EXPERIMENTS

In experiments, researchers deliberately introduce a certain describable set of circumstances and then note what unique effects those circumstances are observed to have. The circumstances are introduced into a carefully controlled context of other circumstances. Those other, background circumstances provide the context for what the researcher is studying; those background circumstances determine the boundary conditions of whatever causal connections the experiment can be taken to establish. Of course, they also must be held constant across both the experimental and control groups.

Clearly, then, conclusions drawn from experiments seem quite restricted. However, researchers are very rarely interested in drawing conclusions only about the circumstances they studied. In fact, the typical impetus for studying a certain set of conditions is interest in a broader set of conditions the studied one is thought to represent. Consider the El-Bassel et al. study: Clearly the purpose of the research was not simply to show that in that particular prison, those particular women selected for treatment had changed for the better.

As the literature review suggests, women, especially drug-using and incarcerated women, are at high risk of dying from AIDS and there is "mounting evidence . . . that increasing AIDS knowledge by itself will not lead to significant behavioral changes". Thus the study evidence is meant to contribute to a wider discussion of what works in the prevention of HIV disease, not just to assist women in one particular prison at one point in time.

One important basis for the credibility of any study, including experiments, is the theory that is used to explain the results. Efforts to enhance internal validity in experiments helps to rule out alternative explanations for the findings. However, the general credibility of the original hypothesis and the theory on which it is based is important as well. Experimental findings lend credibility to a theory by showing that the theory can predict certain observations and events. However, it is rare in the psychological, social, and medical sciences to encounter experimental findings that are compatible with only one theory. Therefore the credibility of an experimental study must also rely on evaluations of theory that are external to the study itself.

The Importance of Replication

The generality of any study's conclusions, including the ones that can be drawn from an experiment, has been termed its external validity. In terms of the traditional framework for understanding causality, externally valid experiments are ones in which the inferred causal connections are also valid under conditions external to those studied. No single study, no matter how well it is designed and how carefully it is executed, can be assumed to have complete external validity.

The science of the helping professions, like all other sciences, has precious few universal laws, that is, generalizations, theories, or propositions that have been shown to hold true under any and all conditions and circumstances. However, the external validity of a study is typically enhanced by including conditions in it that fairly represent circumstances outside of it.

The external validity of a conclusion is most convincingly enhanced by demonstrating it again. This repeated demonstration is called a replication. It refers quite simply to conducting the same study again. Because repeated demonstrations and observations are necessarily conducted under different conditions, each successive replication provides more reason to believe that the "effect" observed will reoccur following the manipulation under still other sets of different conditions.

The enhanced belief afforded by replication in an effect's wider applicability is greater to the extent that successive replications are independent and different from one another, that is, conducted by different experimenters, in different geographical locations, or using somewhat different measures. If similar results are obtained across all the variations—especially if those variations have been systematically selected so as to be deliberately wide over the most relevant circumstantial variables, as is the case in large-scale, national, multisite collaborative studies typically sponsored by federal agencies such as NIMH—the conclusion drawn from them can be held with a great deal of confidence.

It should also be noted that replication is important in all forms of research, not just in experiments. However, some forms of research, such as descriptive and relational studies, typically can employ larger and more varied samples than experiments can, which adds to the credibility of the generalizations made from them. Fixed method studies in which samples are small, such as in experiments and single-subject designs, thus are more often replicated. Unfortunately, however, replication research is not conducted or published as often as it should be, which is a serious problem for knowledge development in general.

THE STRENGTHS AND WEAKNESSES OF GROUP EXPERIMENTS

Experiments are the method of choice when a variable of interest can be manipulated at will by the experimenter and when the research question asks about whether one or more manipulatable variables can produce change in one or more prespecified effects. Under these conditions, well-designed and well-executed experiments can offer highly credible answers to the specific question, including information about treatment techniques, service delivery, and social programmes.

Because experiments have traditionally been considered the pinnacle of scientific method in the social sciences, the rules of method for the design and conduct of experimental studies have been described in great detail. In particular, since experiments aim to demonstrate a logical connection between one event and another, a great deal of attention has been paid to designing experiments in a way that will support such a conclusion.

Experimental studies are therefore often quite elaborate in method and demanding to conduct, especially in field settings. Experiments have limitations.

They can be applied only to phenomena that the experimenter can completely control. Moreover, since they rest on comparisons, they can be applied only to outcomes that predictably and generally occur. With group experiments, questions are answered by examining what happens on average under certain conditions, rather than what happens in any one individual case.

While an unimportant feature for many purposes, this characteristic of experiments can become a real limitation when the primary interest is in generating knowledge about how to best work with particular people or cases. Such questions are best examined with the kind of research introduced in the next chapter: single-subject designs.

AN EXAMPLE OF EXPERIMENTAL RESEARCH

The El-Bassel et al. study of an intervention to assist drug-using women in prison reduce their risk of contracting HIV disease illustrates the usefulness of the experimental method. As the HIV epidemic continues in the United States, increasing numbers of women, especially young women, have been diagnosed with AIDS and ways to help slow the spread of HIV disease among women are urgently needed. The study's literature review cites studies showing that women in prison in particular are at high risk of HIV infection.

In addition, the literature review is used to show that both interpersonal skills and social support can reduce sexual risk-taking, although there were at that time no studies examining the use of these techniques to help drug-using women in prison.

The researchers wanted to know whether or not a programme designed to enhance these skills—interpersonal skills and social support—would or would not help these women in reducing their risk-related behaviour after release. The idea is that information alone is not enough; people need to know how to put that information effectively to use and to get support for doing so.

The experiment was organized to examine a specific hypothesis: ". . . that women in the [treatment] group would exhibit higher rates of safer sex practices, perceived vulnerability to HIV/AIDS, improvement in coping skills, perceived emotional support, and sexual self-efficacy at follow-up one month after release" than women receiving a more limited intervention. According to the logic of experiments, then, the research design must allow for a comparison of the "outcome" phenomena or dependent variables under two conditions: in this case with a "standard" information-based treatment or with an experimental one that also aimed to enhance interpersonal skills and social support. The nature of the treatment received must be controlled by the researchers, and the way the participants in both groups are observed must be the same or at least equivalent.

The outcome of an experiment is a statement about whether or not the hypothesis was supported by the evidence. The intervention studied, the group designed to reduce high-risk behaviours, was carefully defined. A detailed description of the groups and how they were conducted is included in the study report. The description emphasizes how the groups addressed the enhancement of interpersonal skills and social supports specifically. This relationship of the intervention studied to the concepts selected as important in theory, and to the outcomes assessed is critical to any analytic generalization that can be made from such a study.

The logic of experimental design also requires that the groups being compared be the same to begin with so that any differences observed after treatment can logically be attributed to the intervention received. This requirement is generally addressed by using random assignment. Indeed, in the El-Bassel study all the women in one particular prison who met study criteria and who consented to participate were randomly assigned to the experimental group, which offered 16 skill-building sessions, or a "control" group, a 3-session AIDS information group. In this way, differing motivations to undertake a more or less demanding treatment were ruled out as an explanation for the results. The sample in the El-Bassel et al. study was characteristically specialized.

First of all, they were women in prison, who certainly are not statistically representative of all women at risk for HIV infection. In addition, only some women in the prison were eligible to participate; they had to be "between 18 and 55 years of age, convicted and serving a sentence between three months and one year, scheduled for release within 10 weeks [to certain parts of New York City], and self reported drug users . . .". These criteria were designed to involve women in the study who had a history suggesting risk, who could consent to participate in the research for themselves, and who would soon be in a position to deal with high-risk sexual situations in the community. Thus it was a sample in which the phenomena of interest—changes in risk-related behaviour—could best be observed.

In the El-Bassel et al. experiment, several distinct outcomes involving both behaviour and attitudes were assessed. In addition, one of the outcome, or dependent, variables, coping skills, was measured in a specific, predefined way using a previously developed instrument, a Coping Skills Questionnaire.

The others were based on answers to self-report questions that were quantified in specified ways. Data were collected and analyzed in the same ways in both groups of study participants—those who did and did not get the experimental intervention.

As is common in experiments, multivariate statistical analyses were conducted, in this case logistic regression analysis, a form of multiple regression analysis that allows for independent variables that are categorical, like marital status and treatment received, as well as continuous, like the number of sessions of treatment attended.

The authors state that their hypothesis received "some support" because only some of the outcome variables—safer sex behaviour, coping skills, and perceived emotional support—were shown to differ according to whether or not the participants had received the experimental treatment. The fact that no difference was found in AIDS knowledge was not surprising since both groups, experimental and control, had been exposed to this information.

The fact that safer sex behaviours changed as a result of the group treatment although perceptions of risk did not "seem[s] to challenge a central tenet of the health belief model," the authors concluded. Thus both practical and theoretical implications of the results are considered. As the El-Bassel et al. study illustrates, experimental designs are logically simple but complex to conduct. Retention of study participants was a challenge. Nevertheless it illustrates well the logic of experimental design and the standards traditionally used to judge how well any experiment has satisfied the logical conditions that lead to confidence in the conclusions drawn.

THE LOGICAL VALIDITY OF EXPERIMENTAL DESIGNS

Because so much attention has been paid to the experiment as a method of empirical research, there is a great deal of special terminology used in the discussion of experimental designs. The experimenter's theoretically derived statement concerning a change-result relationship, or the presumed "cause-effect" relationship, is called a hypothesis. More precisely, it is called a hypothesis of difference, because it is a hypothesis that the presence or absence of the presumed cause will make a difference in the occurrence or nonoccurrence of the presumed effect, which is determined by comparing outcomes in two groups.

Hypotheses of difference, then, make theoretically based statements about links between a manipulation, the presumed "cause," and a predefined observed change, the presumed "effect." The manipulated variable is known as the independent variable; it is a variable, since its values can vary, and it is independent, presumably, of everything except the experimenter's decision about whether it will be absent or present and to what degree. The phenomenon expected to change is known as the dependent variable; it is a variable, again, since its values can vary, and it is called dependent to signify that it presumably depends on the independent variable for its presence or absence. In the El-Bassel et al. study, the independent variable is the intervention received; the dependent variables were safer sex behaviour, coping skills, perceived availability of emotional support, sexual self-efficacy, perceived vulnerability to HIV/AIDS, and AIDS knowledge. Clients who did and who did not receive the experimental treatment were compared on the dependent variables holding all other things equal—in Latin, ceteris parabus.

As mentioned earlier, ceteris parabus is an important qualifier. The comparison between the two conditions is valid only to the extent that all other things, that is, all factors other than the deliberately varied independent variable, were in fact equal. If there were other inequalities, differences observed in the dependent variable might result from those other differences rather than from the arranged difference in the independent variable. Recalling Mill's logic, the other differences become rival hypotheses explaining observed differences in the dependent variable that cannot logically be ruled out. Therefore, inequalities between the compared conditions compromise the value of an experiment for establishing a link between an independent and a dependent variable. The coexisting, unintended inequalities are called confounds because they are mixed up, or confounded, with the independent variable.

Confounds threaten what is known as the internal validity of an experiment. An experiment's internal validity refers to the extent to which it is legitimate to attribute observed dependent variable differences unambiguously to independent variable differences. Internally valid experiments permit concluding unequivocally, following the logic of Hume, Mill, and Popper, that the independent variable and dependent variable are linked. When the design of an experiment lacks internal validity, it is impossible to determine whether dependent variable differences were caused by variations in the independent variable or whether they were caused by the other factors. A hypothetical example will serve to illustrate this point. Suppose a researcher believes that therapy A is more effective than therapy B for the treatment of depression. Testing this hypothesis would involve dividing a group of depressed clients into two equivalent subgroups, administering therapy A to one of the subgroups of depressed clients, administering therapy B to the other subgroup of depressed clients, and then comparing the average depression levels in the two subgroups posttreatment.

The experimenter hopes that there will be a specific posttreatment difference in the dependent variable, namely that the group of clients who received therapy A will be less depressed than the group of clients who received therapy B. If this difference in the dependent variable is, in fact, observed, the experimenter would like to attribute it to the deliberate difference in the independent variable.

Now suppose for a moment that the two therapies were administered by different therapists. And suppose that the therapists for therapy A were well respected, seasoned clinicians while the therapists for therapy B were barely trained but well-intended undergraduate paraprofessionals. Clearly, under these circumstances it would not make sense to suggest that the dependent variable difference could be unambiguously attributed to the intended independent variable difference. Since there were two major differences in the treatment given the two groups, that is, the type of therapy and the experience of therapists, it is not possible to determine whether it was the first or second of them, or the combination, that caused the posttreatment dependent variable difference.

In general, this same ambiguity is found whenever an additional variable changes along with the intentionally altered independent variable. The ravages confounds introduce into experiments, in terms of weakening the legitimacy of the independent-dependent variable link that can be made on the basis of them, are profound. The impact of confounds is so great, in fact, that not having them is considered by some methodologists to be the single most important criterion of a good experiment.

Confounds in Experimental Design

In a famous 1963 monograph, Campbell and Stanley provided an in-depth discussion of the confounds encountered in experimental research. Each of these confounds, discussed below, is a difference that varies alongside the intended difference in the independent variable. When present, each undercuts the internal validity of an experimental design by permitting an alternative explanation, or rival hypothesis, other than the independent variable difference, for why dependent variable differences might be observed. These common confounds are most easily understood through example. Suppose that El-Bassel and colleagues had only enrolled women in the experimental treatment to enhance social supports and interpersonal skills and then studied them after release, finding that the women had gained in social support and other measures at the second assessment point. Notice that this study strategy involves making a before-and-after comparison using only one group of people.

This type of design is often called pre-post. Pre-post designs are not internally valid because the pre- and postconditions are not identical except for the occurrence of the independent variable. We will now consider why this is the case— what may be different in addition to the presence versus absence of the independent variable and therefore what else might be explaining any dependent variable differences observed before and after intervention.

History

The first rival hypothesis, introduced inherently with the passage of time, is termed history. These other events, whether they are known or not, represent confounds, and the passage of time always allows for the occurrence of additional historical factors that could account for observed pre-post independent variable effects. For example, a major public education campaign about HIV disease might have taken place in the home community to which the women had been released. The street supply of an illegal drug commonly used by the women may have been reduced, resulting in fewer occasions for intoxication and unsafe sexual encounters. Such developments might provide alternative explanations for any positive changes in outcome observed.

Maturation

Campbell and Stanley termed the second rival hypothesis maturation. Linguistically, maturation refers to the process of becoming older; in psychology it also refers to the emergence of personal and behavioral characteristics through processes of experience and growth. In the example of the drug-using women in prison, it may be that the process of conviction, imprisonment, and release alone helped move them along on average to better functioning. In addition, there is evidence that some people "age out" of addiction. These factors rather than the experimental treatment could be alternative explanations for posttreatment change arising from within the individuals involved. Again, with the passage of time, there is always the possibility that some naturally unfolding developmental process might account for observed differences; maturation becomes a rival hypothesis to explain pre-post gains.

Repeated Testing

Whenever independent variable effects are monitored through pre- and posttesting, the effects of repeated testing become a potential confound. In the El-Bassel et al. example, women were asked questions about themselves and their behaviour before and after they had treatment. It is possible on hearing the same questions a second time that the study participants consciously or unconsciously gave more positive answers about themselves because they had time to reflect on what the researchers might think about their answers. Repeated testing, then, is another confound; like history and maturation, it introduces a rival explanation for apparent pre-post intervention changes in the dependent variable.

Instrumentation

The fourth common confound Campbell and Stanley discussed is instrumentation. This term refers to score shifts that result from changes in the measure rather than changes in the measured: Scores look different because the measuring instrument has changed. In the El-Bassel et al. study, for example, the same interviewers were used at all data collection points. It is possible that the interviewers, once they got to know the women, consciously or unconsciously "helped" the women they interviewed give more positive answers at follow-up because they had come to care about them. Where instrumentation changes are a possibility, it is not possible to determine whether the independent variable or the confound resulted in the higher spring grades.

Regression to the Mean

The fifth common confound is called regression to the mean. This confound is easy to understand intuitively, despite the fact that it is a statistical concept. Scores on any measure of a phenomenon, including behavioral self-report measures, are unfortunately imperfect: All scores are comprised of some truth and some degree of error. This error component, when not due to systematic bias, is assumed to be randomly distributed. Therefore some of the scores at either extreme, that is, some of the scores that are extremely low and some of the scores that are extremely high, will be extreme partially because of the influence of error. Upon retesting, that error will be more likely to cancel itself out than it will be to reoccur; the result of this fact is that on retesting extreme scorers are more likely to get less extreme scores than they are to get equally or more extreme scores.

As applied to our example, regression to the mean might apply. Because women's self-reported behaviour was being compared to what it had been in the month before imprisonment, their posttreatment behaviour was being compared to the point in their lives when it had been the worst, the point at which they got into trouble with the law. Extreme scores always tend to normalize, that is, to move or regress towards the average score, called the mean, upon retesting. In the case of a pretest, a score's regressing towards the mean would ordinarily be taken as indicating improvement. Therefore, whenever working with those most in need, which we often do as helping professionals, unless we take special design precautions, it is not possible to attribute improved scores to the treatment rather than to regression to the mean.

The Use of a Control Group

Fortunately, in experiments special design precautions can be introduced to neutralize the possible effects of regression to the mean and the other potential confounds just described. Without question, the most important special design element that can be introduced into an experiment is what is known as a control group. Control groups provide a basis for comparison. Theoretically, including a control group in an experiment should provide a basis for a pure, that is, unconfounded, comparison between what happens when the independent variable is introduced in one way versus what happens when the independent variable is not introduced in exactly that same way. In fact, the use of a control group can eliminate all of the threats to the internal validity of an experiment so far discussed.

The El-Bassel et al. design involving two equivalent groups of women studied during the same period of time eliminates history as a confound. If events other than the experimental treatment that might affect the dependent variables occurred during the study period, they would be likely to affect both groups; thus if differences between the groups had been found, it would have been hard to argue that historical events had affected one group but not the other. Similarly, maturational processes can be expected to have affected both groups. In addition, the fact that both groups of women—those who got the experimental and those who got the standard treatment—completed the same assessment controls for the effects of testing and of instrumentation. If testing or instrumentation effects were occurring, they should have affected both experimental and control groups equally. On the other hand, a finding that two groups that were the same before the intervention but different after it would suggest that testing or instrumentation did not create the difference—if they both underwent the same testing procedures. As to regression to the mean, the inclusion of two equivalent groups of "low scorers," that is, of women in trouble, eliminates regression to the mean as a competing explanation for the results.

These precautions that preserve internal validity do not eliminate possible effects from these confounds; what they do is ensure that all study participants remain equally susceptible to them. This equal susceptibility means that experiments really make the following comparison: what happens when the independent variable is introduced under a set of conditions that include some factors in addition to the independent variable that may have an effect on the dependent variable, versus what happens when the independent variable is not introduced under a set of conditions that include the same factors other than the independent variable that may have an effect on the dependent variable. Experiments demonstrate that independent variables make differences above and beyond whatever differences result from other, confounding factors.

This requirement, that is, that experiments must make comparisons between effects observed under conditions equally susceptible to effects from confounds, includes both a freedom and a constraint for the researcher. The freedom is found in the fact that it is not necessary to be able to arrange a sterile, static environment to demonstrate independent variable effects.

It is hard to imagine how experimental design would have any relevance to the helping professions without this freedom since many, if not most, of the most pressing questions facing social workers and other human service professionals can be meaningfully asked only in the ever-changing and certainly not sterile world of practice, rather than in the static, sterile conditions sometimes possible in the laboratory. Balanced against this freedom is a constraint: Whatever the factors are that were part of the context when the effects of the independent variable were demonstrated become potential boundary conditions, that is, conditions that limit when the independent variable can be expected to exert those same effects.

These boundary conditions include what have traditionally been called confounds affecting external validity due to the interaction of the treatment or manipulation itself and the process of study. So, in the case of the El-Bassel et al. study, for example, this single experiment cannot "prove" that the same effect of the experimental treatment would be found in the absence of participation in the other pre- and posttesting that took place as part of the study or if the selection of participants had been made from some entirely different group. However, its internal logic, or validity, is clear.

It does not detract from the value of experiments, or any form of research, to recognize that most scientific observations hold true only under certain conditions. What matters is recognizing what those boundary conditions are so that effects will not be anticipated in the absence of them. What is essential here is to note that this problem of boundary conditions affects all kinds of research and not just experiments and that theoretical logic and independent replications of observations have always been acknowledged to be essential for generalization from experiments and other forms of research.

Other Forms of Control in Experiments

The term control actually has several meanings in experimental design. All are variants of the idea of allowing a valid comparison, that is, a comparison uncontaminated by effects from extraneous variables—from variables that are not of primary interest in the study. For example, when considering control groups, the term control means equally susceptible to the effects of the kinds of confounds that were described above. Sometimes experimenters will speak of "controlling for" a variable when they mean to prevent a variable from having an effect. This variant of control is analogous to the type of precaution exercised in the legal system when a judge sequesters a jury to prevent media accounts and others not part of the jury from being able to exert an influence on the jury's decision.

Sometimes experimenters arrange an experiment in such a way as to prevent a potentially influential variable from having an effect on the participants under study and their responses. When this is done, the experimenter is said to have "controlled for" the effects of that excluded variable. Persons not very familiar with research often believe this second meaning of control is the only meaning researchers have for the term. That is, lay people usually assume that to control for a variable always means to eliminate its effects.

This elimination is one meaning of the term control and in fact is the meaning that justifies asking questions about human behaviour in the laboratory rather than in the “real world,” since certain variables can only be eliminated in the laboratory, where the experimenter has more control over the situation.

However, it is only one meaning of the term, and it certainly is not the most fundamental one. There is a third way in which the term control is used in experimental design. Sometimes it is clear that an extraneous variable has had an effect on some research participants and not on others, and that different numbers of participants were affected in the experimental than in the control group.

In this circumstance, the extraneous variable is confounded with the independent one. That is, participants in the two groups were different in two ways at once. First, they were different with respect to whether or not they were administered the independent variable. Second, they were different by virtue of being affected by some extraneous variable: More of the people in one of the groups rather than the other were affected by it, so that the groups were not equally susceptible, or exposed to, the effects of that extraneous variable.

It is possible to control for the effects of such an extraneous variable in a variety of ways, including through multivariate statistical analysis like that used in the El-Bassel et al. study. For example, marital status and the number of treatment sessions attended were examined along with treatment group participation when examining the outcome of maintaining safer sex practices. This analytic strategy controls for the effects of those potentially confounding variables, rendering the main comparison ceteris parabus.

The Functions of Control Groups

The notion of control, then, implies making comparisons ceteris parabus, which is usually done by including an appropriate control group in an experimental design. Useful control groups are ones that are as susceptible to the influences of all variables other than the independent one as is the experimental group. This equal susceptibility allows study of the effects uniquely attributable to the presence of the independent variable.

The comparison between experimental and control groups is useful only when the conditions described in the Latin phrase ceteris parabus are met, that is, only when all other things are equal. If ceteris parabus, then it will be permissible to use the logic of experimental reasoning to argue that observed differences between the experimental and control groups in what happens are definitively attributable to the experimenter-manipulated independent variable.

Although the inclusion of a control group can help strengthen the internal logic of an experimental design, there are factors that can render a control group not useful. In essence these factors are ones that destroy the fundamental comparability of the experimental and control groups. This fundamental comparability can be destroyed before, during, or after the occurrence of the independent variable. The way participants in an experiment are assigned to experimental and control groups within a study turns out to be essential to making them useful and to avoiding some additional potential confounds to the logic of the design.

Random Assignment

The method of determining group assignment that has the best chance of ensuring comparability between the experimental and control group is just that: chance. Chance, or pure randomness, is ultimately indifferent; if indifference is used as the basis for assigning participants to groups, then chances are that nothing will favour one group over the other, which is what results in inequality.

Using chance to assign participants to groups is of fundamental importance in working towards ceteris parabus, which is of course the condition that legitimizes attributing observed differences to the manipulated independent variable. There are several methods for achieving random assignment that are discussed in chapter 10. All of them employ some unbiased, chance-based method, like drawing names or numbers from a container, as in a lottery, or using tables of random numbers.

When random assignment, either simple or stratified, is used, several additional potential confounds are either eliminated or minimized. The first of these is called selection. Selection is a confound that arises when some procedure other than chance is used to assign participants to groups and that procedure results in initial group differences. Initial group differences, as has been mentioned many times, preclude arguing that observed after-the-fact differences arose because of the treatment rather than because of the initial group difference or some correlate of it.

Selection Effects

Studies confounded by selection effects are ones in which the experimental and control groups are not comparable from the outset by virtue of the fact that they were selected using different criteria. As an example, suppose an administrator wants to evaluate the effects of supervising student interns on social worker morale. Experimental design would imply conducting that evaluation by comparing the morale of social workers who are supervising student interns against the morale of social workers who are not supervising student interns.

Now, suppose the administrator identifies the workers who will be given student interns by asking for volunteers. Is there not reason to believe that the workers who volunteer to supervise students might systematically differ in important ways from those who do not volunteer? This preexisting difference would be just as likely to explain any differences in morale after supervising as would the presence or absence of the student supervise.

It is not possible to untangle the factors that would be responsible for any dependent variable differences or whether dependent variable differences were attributable to the joint effects of the preexisting difference—the willingness to volunteer—and the impact of the independent variable. The comparison is not ceteris parabus: All other things, that is, things other than the independent variable, were not equal since the groups were different both in terms of the independent variable, the presence or absence of the student intern, as well as whatever other differences preexisted between the group of workers who had volunteered for the supervision versus the group of workers who had declined to volunteer.

It may seem that the best way to arrange comparability between experimental and control groups is to work at it. Put differently, it may seem that the best way to ensure that the experimental and control groups are equal at the outset is to study all the eligible participants and then deliberately assign some of them to the experimental group and others to the control group, using care to make sure that overall, the two groups are well matched. And in fact, sometimes this is the best strategy: If there is one very important known variable that must be equal between the two groups before the treatment is applied, stratified random assignment can be used, as described in the section on sampling below. In general, however, while matching may seem a logical strategy, very rarely can an experimenter identify before the study all the variables important that groups are equivalent on. Also, even when these variables can be identified, usually there is more than one so that identifying matched participants becomes quite cumbersome. Nevertheless, whether an important matching variable can or cannot be identified, the question remains of how to determine whether each subject should be assigned to the experimental or the control group.

It turns out that the best way to ensure original, before-the-fact comparability between an experimental and a control group is not matching. Even the best-intended researcher cannot be fair: Biases in group assignment always enter in, whether the researcher is or is not aware of them. It also turns out that there is no way to guarantee that experimental and control groups will be equal before the fact: There is always a chance that, no matter what precautions are taken, they will be different in some important known or unknown way.

Differential Attrition

Random assignment to groups has been said to provide some protection against another confound, although random assignment by no means affords the same level of protection against this confound that it does against selection. This confound is termed differential attrition. Attrition refers to loss of participants during the conduct of the research. Attrition always presents a threat to the generalizability of a study in that what is observed can be considered generalizable only to the sort of people represented by the participants who remained in the study. With differential attrition, the degree of participant loss is different between the treatment and control conditions: More participants are lost in one of the conditions than in the other.

As a result, ceteris parabus is lost: The two groups are no longer equal because of the differential loss, so comparing them is not appropriate. The El-Bassel et al. study discusses this problem and how it was handled on page 135. Consider also the following example. Suppose a psychiatrist wants to compare the effectiveness of a traditional drug to the effectiveness of a new drug for the treatment of depression. And suppose that the new drug turns out to have many more unpleasant side effects than does the traditional drug.

It is quite likely that under these circumstances many more of the patients receiving the new drug will drop out of the study. If this happens, the end comparison will be between patients receiving the old treatment and the few patients receiving the new treatment that tolerated the drug well enough to remain in the study. The fact that more patients in the treatment group dropped out during treatment renders the two groups noncomparable after treatment, and this would be the case even if random assignment had been used to initially assign participants to groups. So while random assignment provides some protection against differential attrition, it cannot prevent it.

Verifying the Independent Variable

A final issue in drawing valid conclusions from experimental studies has to do with verifying the independent variable. El-Bassel et al. do this in two ways. First of all they describe the nature of the experimental intervention in some detail. Omitting from the study this information about the independent variable makes an experiment difficult to replicate. This leads to what is called the "black box" approach, in which the content of the treatment that proved effective remains an unknown. Verifying the independent variable also means making some assessment of its actual use as part of an experiment. El-Bassel et al., for example, recorded the attendance of the women in both the treatment and control groups. This information is termed dosage, that is, the amount of exposure to the intervention of interest. However, the analyses they reported did not show a statistically significant relationship between the number of treatment sessions attended and outcome. The conceptual link between intervention and outcome is strengthened when the amount of exposure to a treatment is shown to have an impact on outcome.

5

Techniques of Data Collection

Basic requirements for scientific data are that it should be reliable and impartial. In Sociology these conditions are hard to meet. Yet numerous methods are used to minimize errors in data. Some of the commonly used sources in collecting data are:

- Existing materials including the official statistical record and historical and contemporary documents.
- Social surveys through questionnaire and schedules
- Interviewing
- Observation-Participants and non-participant.

EXISTING MATERIAL

Statistical Sources: Government statistics particularly census or statistics produced by large industrial or commercial firms, trade unions or other organizations provide one important account of data which sociologist can use in their analysis.

An outstanding example of the imaginative use of official statistics in the positivist tradition is the study of suicide made by the famous French sociologist Emile Durkheim in the 19th century.

However official statistics are the kind of data that are not collected by sociologists themselves and so there problems while analysing the data.

Historical documents: Records and accounts of qualitative kind for example relating to belief, values, social relationship or social behaviour may also be contemporary or may refer to earlier periods. There are several difficulties immediately present themselves in the use of records from the past.

Few chroniclers of social relation and social action record observations in the systematic way in which the sociologists are interested. There are often intriguing and sympathetic records but the information that is vital to the sociologist is often missing.

Contemporary Records: Contemporary records relating to social relationship and social behaviour are seldom used as the sole source of information and sociological research. They are usually one source of a particular account or achievement.

SOCIAL SURVEY

The basic procedure in survey is that people are asked a number of questions on that aspect of behaviour which the sociologist is interested in. A number of people carefully selected so that their representation of their population being studied are asked to answer exactly the same question so that the replies to different categories of respondents may be examined for differences.

One type of survey relies on contacting the respondents by letter and asking them to complete the questionnaire themselves before returning it. These are called Mail questionnaires.

Sometimes questionnaires are not completed by individuals separately but by people in a group under the direct supervision of the research worker. A variation of the procedure can be that a trained interviewer asks the questions and records the responses on a schedule from each respondent.

These alternate procedures have different advantages and disadvantages. Mail questionnaires are relatively cheap and can be used to contact respondents who are scattered over a wide area.

But at the same time the proportion of people who return questionnaires sent through post is usually rather small. The questions asked in main questionnaires have also to be very carefully worded in order to avoid ambiguity since the respondents cannot ask to have questions clarified for them.

Using groups to complete questionnaires means that the return rate is good and that information is assembled quickly and fairly. Administrating the interview schedules to the respondents individually is probably the most reliable method. Several trained interviewers may be employed to contact specific individuals. The questionnaires and schedules can consist of both close-ended and open-ended questions. Also a special attention needs to be paid to ensure that the questionnaires are filled in logical order.

Where aptitude questions are included great care must be exercised to ensure the proper words are used. In case of schedules emphasis and interactions may also be standardized between different individuals and from respondents to respondents.

Finally proper sampling techniques must be used to ensure that the sample under study represents the universe of study. In order to enhance the reliability of data collected through questionnaires and schedules, these questionnaires and schedules must be pretested through pilot studies.

INTERVIEWING

Social surveys may depend either on questionnaires that are self-administered or on schedules completed by trained research workers personally interviewing then is not a method of data collection distinct from social surveying but rather a technique which may vary from the brief formal contact as when the interviewer is working for the firms public opinion consultants or a market research organization and simply asks a housewife a few highly specific questions on limited range of topics to a long interview in which the research worker allows the respondents to develop points at leisure and take up others as he chooses.

The brief formal interview in which the working of the questions and the order in which they are asked is fixed is called structured interview while the freer discursive interview is called unstructured interview.

The object of using structured interview is to standardize the interview as much as possible and thus to reduce the effect that the interviewer's personal approach or biases may have upon the result and even when structured interviews are used, proper training can do a lot to ensure further the reliability and validity of research. The personality of the interviewer and the social characteristics that the respondents attribute to him can be having influence on the result.

The effort of interviewer's bias can be estimated by comparing one interviewer's result with other. The problem of interviewer's bias in an unstructured interview is much greater. Here the interviewer is left to his common devices as far as the way he approaches a respondent is concerned. There is no fixed list of questions to work through. Instead the interviewer may work from a guide that will remind him of the topics he wishes to cover. The training of the interviewer is crucial here not simply training in the social skills of keeping the conversation going on a topic that the respondent may not be very interested in but also in acquiring sensitivity to those things his respondents tells him which are specially relevant to the theoretical topics he is pursuing. This means that unstructured interviews can be carried out by people trained in sociological theory. They are then able to size upon stray comments made by the respondents which can be developed and lead on to important theoretical insight.

OBSERVATION: PARTICIPANT AND NON PARTICIPANT

The rationale behind the use of observation in sociological research is that the sociologist should become party to a set of social actions sufficiently able to be able to assess directly the social relationship involved. The degree of involvement may vary considerable from being merely a watcher on the sidelines to be deeply involved in and being a part of what is going on. The former type of observation techniques are called non-participant while the latter is called participant observation. Sometimes one way observations screen have been used to watch groups in actions that they are unaware that they are being watched and the observer cannot affect their actions by his presence. The sociologist is visibly present and is a part of the situation either as a sociologist or in another guise. Where the sociologist is merely an observer it is usually assumed that he knows enough about what the actors are doing to be able to understand their behaviour. Any sociological observer has then to some extent be a participant observer he must at least share sufficient cultural background with the actors to be able to construe their behaviour meaningfully but the degree of participation and of sharing of meaning may vary considerably. Examples of such studies are Nel Anderson's study of Hobo-Indians and William White study of Street Corner Society.

SAMPLING

For practical and cost reasons, it is often impossible to collect information about the entire population of people or things in which social researchers are interested. In these cases, a sample of the total is selected for study. Most statistical studies are based on samples and not on complete enumerations of all the relevant data. The main criteria when sampling are to ensure that a sample provides a faithful representation of the totality from which it is selected, and to know as precisely as possible the probability that a sample is reliable in this way. Randomization meets these criteria, because it protects against bias in the selection process and also provides a basis on which to apply statistical distribution theory that allows an estimate to be made of the probability that conclusions drawn from the sample are correct. A statistical sample is a miniature picture or cross-section of the entire group or aggregate from which the sample is taken. The entire group from which a sample is chosen is known as the population, universe or supply.

SIMPLE RANDOM SAMPLING

The basic type of random sample is known as a simple random sample, one in which each person or item has an equal chance of being chosen. Often a population contains various distinct groups or strata that differ on the attribute that is being researched.

STRATIFIED RANDOM SAMPLING

Stratified random sampling involves sampling of each stratum separately. This increases precision, or reduces time, effort and cost of allowing smaller sample sizes for a given level of precision. For example, poverty is known to be most common among the elderly, the unemployed and single parent families, so research on the effect of poverty might will sample separately each of these three strata as part of a survey of poverty in the population as a whole which would permit the total sample size to be reduced because the investigator would know that the groups most affected by poverty were guaranteed inclusion.

CLUSTER SAMPLING

Cluster sampling is sometimes used when the population naturally congregates into clusters. For example, managers are clustered in organizations, so a sample of managers could be obtained by taking a random sample of organizations and investigating the managers in each of these. Interviewing or observing managers on this basis would be cheaper and easier than using a simple random sample of managers scattered across all organizations in the country. This is usually less precise than a simple random sample of the same size, but in practice the reduction in cost per element more than compensated for the decrease in precision.

MULTI-STAGE SAMPLING

Sampling may be done as one process or in stages, known as multi-stage sampling. Multi-stage designs are common when populations are widely dispersed. Thus a survey of business managers might proceed by selecting a sample of corporations as first stage units, perhaps choosing these corporations with a probability proportionate to their size, and then selecting a sample of managers within these corporations at the second stage. Alternatively, a sample of individual factories or office buildings within each corporation could be chosen as the second stage units, followed by sample of managers in each of these as a third stage. Stratification can also be used in the design, if for example occupational sub-groups are known to differ from each other, by selecting state such as personnel, production, and finance management and sampling within each of these. For sampling to be representative, one needs a complete and accurate list of the first stage units that make up the relevant population, a basic requirement that is not always easily met. This forms the sampling frame. Selection from the frame is best done by numbering the items and using a table of random numbers to identify which items form the sample, though a quasi-random method of simply taking every item from the list is often appropriate. The reliability of a sample taken from a population can be assessed by the spread of the sampling distribution, measured by the standard deviation of this distribution, called the standard error. As a general rule, the larger is the size of the sample the smaller the standard error.

AREA SAMPLING

In sampling of this kind small areas are designated as sampling units and the households interviewed include all or a specified fraction of those found in a canvass of these designated small areas. The basic sampling units or segments chosen may be relatively large or relatively small depending on such factors as the type of area being studied, population distribution, the availability of suitable maps and other information and the nature and desired accuracy of the data being collected.

DATA COLLECTION

Data Collection helps your team to assess the health of your process. To do so, you must identify the key quality characteristics you will measure, how you will measure them, and what you will do with the data you collect.

What exactly is a key quality characteristic? It is a characteristic of the product or service produced by a process that customers have determined is important to them. Key quality characteristics are such things as the speed of delivery of a service, the finish on a set of stainless steel shelves, the precision with which an electronic component is calibrated, or the effectiveness of an administrative response to a tasking by higher authority. Every product or service has multiple key quality characteristics. When you are selecting processes to improve, you need to find out the processes, or process steps, that produce the characteristics your customers perceive as important to product quality.

Data Collection is nothing more than planning for and obtaining useful information on key quality characteristics produced by your process. However, simply collecting data does not ensure that you will obtain relevant or specific enough data to tell you what is occurring in your process. The key issue is not: How do we collect data? Rather, it is: How do we obtain useful data?

WHY DO WE NEED TO COLLECT DATA

Every process improvement effort relies on data to provide a factual basis for making decisions throughout the Plan-Do-Check-Act cycle. Data Collection enables a team to formulate and test working assumptions about a process and develop information that will lead to the improvement of the key quality characteristics of the product or service. Data Collection improves your decision-making by helping you focus on objective information about what is happening in the process, rather than subjective opinions. In other words, I think the problem is... becomes... The data indicate the problem is....

WELL-DEFINED DATA COLLECTION PROCESS

For your team to collect data uniformly, you will need to develop a Data Collection plan. The elements of the plan must be clearly and unambiguously defined- operationally defined. You may want to pause here and review the Operational Definitions module before you go on. Why does a team need Operational Definitions in order to collect useful data? Let's say three people are collecting data on the time it takes to perform a certain process step. Unless the exact moment when each action begins and the exact moment when it ends are operationally defined, each data collector will observe and record data based on his or her own understanding of the situation.

The Data Collection process will not be standardized or consistent. You will have collected data, but it probably won't be much good to you. Worse yet, you may make changes to your process based on flawed information.

Data Collection can involve a multitude of decisions by data collectors. When you prepare your Data Collection plan, you should try to eliminate as many subjective choices as possible by operationally defining the parameters needed to do the job correctly. It may be as simple as establishing separate criteria and a specific way to judge when a step begins and when it ends. Your data collectors will then have a standard operating procedure to use during their Data Collection activities.

DEVELOP A DATA COLLECTION PLAN

You should develop your Data Collection plan during the Plan Phase of the Plan-Do- Check-Act (PDCA) cycle. The PDCA cycle provides a framework for you to build an understanding of your process and how to obtain and interpret data that will lead to real process improvement. Although they can be time-consuming, planning sessions are extremely important because this is when you establish the guidance that helps you obtain the right data.

THE DATA COLLECTION PLAN ANSWER

Your team needs to develop the answers to the following questions as the basis for a sound Data Collection plan:

- Why do we want the data? What will we do with the data after we have collected them? The team must decide on a purpose for collecting the data.
- In the Plan Phase, your team should develop a working hypothesis which will serve as a guide to future investigation of the process. This hypothesis is an assumption based on already existing data and observations, such as your process Flowcharts or a Cause-and-Effect Diagram the team has prepared. You develop working assumptions and collect data to determine the process changes that will improve the key quality characteristics of your product or service. Your proposed change should be stated as an "If... then" statement.
- IF we change Step X in our process by doing..., we believe as suggested, THEN improve Y, which is a key quality characteristic of our product or service

This action focuses your team on the specific quality characteristic you want to improve, and sets the stage for where you will collect the data.

- Where will we collect the data? The location where data are collected must be identified clearly. This is not an easy step unless you tackle it from the following perspective:
  - Refer to the Flowcharts which depict both the current ("as is") state of the process and the proposed ("should be") state of the process after it has been modified. Focus on the process steps where the key quality characteristic you are trying to improve is produced.
  - Collect data from these process steps. You must collect data twice. First, you collect baseline data before you make any changes in your process. These baseline data will serve as a yardstick against which to compare the results of the process after changes have been made. Then, you must collect data after the change has been imposed on the process. To compare the before and after process, you will probably want to translate your data into graphic form using a Pareto Chart, Run Chart, or Histogram. The use of these tools is explained in separate modules.

  - Collect data on the key quality characteristic of the product or service at the end of your process. Again, before and after data must be collected. The comparison of before and after data validates whether the change actually improved the output of the process.
- What type of data will we collect? In general, data can be classified into two major types: attribute data and variables data.
  - Attribute data give you counts representing the presence or absence of a characteristic or defect. These counts are based on the occurrence of discrete events. As an example, if you are concerned with timely delivery of parts by your store keepers, you could develop a procedure that would give you a count of the number of supply parts they deliver on time and the number they deliver late (defects). This would give you attribute data, but it would not tell you how late a late delivery actually was. Two factors help determine whether attribute data will be useful:
  - Operational Definitions. You need to operationally define exactly what constitutes a defect. For the data collected in the example above to be useful, you would have to operationally define late. This may be a good time to review the module on Operational Definitions.
  - Area of Opportunity. For counts to be useful, they must come from a well-defined area of opportunity. You obtain a single count, or value, from each sample, or area of opportunity. For example, if you are collecting data on the number of defective bayonets received in each shipment of 200, the area of opportunity is the 200-bayonet shipment. The number of defective bayonets in the shipment gives you one count, or data point.
  - Variables data are based on measurement of a key quality characteristic produced by the process. Such measurements might include length, width, time, weight, or temperature, to name a few. Continuing with the parts delivery example, you could collect variables data by tabulating the time it took to process an incoming supply request from receipt to validation of the National Stock Number (NSN); or the time from validation of the NSN to identification of the stock bin where the part is located; or the time required to post the obligation in the OPTAR Log; or the total time from receipt of the request to delivery of the part. This measurement, time, could be used to determine how timely or late the deliveries were.
- Who will collect the data? Many teams struggle with this question, but the answer is simple: Those closest to the data—the process workers—should collect the data. These people have the best opportunity to record the results. They also know the process best and

can easily detect when problems occur. But remember, the people who are going to collect the data need training on how to do it, and the resources necessary to obtain the information, such as time, paper, pencils, and measurement tools.

- How do we collect the right data? You need to remember that you are collecting data for the purpose of improving the process, not the product it produces. Clearly, you want to collect the data that best describe the situation at hand. If you are going to use the data to make predictions about the performance of the process, you should collect small samples at regular intervals—let's say 4 or 5 units every other hour or each day. Since it is important to collect those 4 or 5 units in a short interval of time, you may want to use consecutive units or every other unit.
- But remember, the cost of obtaining the data, the availability of data, and the consequences of decisions made on the basis of the data should be taken into consideration when determining how much data should be obtained and how frequently it should be collected.

DATA COLLECTION PROBLEMS

Remembering that data form the basis for the effective, unemotional communication without which no process improvement effort can succeed, you need to avoid two significant problems associated with Data Collection.

Problem 1: Failure to establish Operational Definitions. You need to define, not simply identify, the following:

- When and how often you will collect the data
- How you will collect the data
- Units of measurement you will use in collecting the data
- The criteria for defects
- How you will handle multiple defects on single products

If you haven't thought about these issues, your Data Collection process may be doomed from the start. This is especially true when more than one person is collecting data. What is meaningful to one worker might not be to another. You have to take the time to develop adequate, clear-cut definitions, and train each collector to use those definitions.

Problem 2: Adding bias to the Data Collection process. You can never eliminate bias, but it is important to minimize it. Here are some ways your data can be biased:

- The process of collecting the data may affect the process being studied. If you are trying to make a process faster, taking data may either speed it up or slow it down.
  - On the one hand, the workers may speed up the way they work in the process, thus skewing the data in their favour. This may occur if they have a perception that the variables data they are collecting will show that they could be more efficient, productive, or effective. Once the Data Collection effort ceases, they may return to their old pace of operations.

– On the other hand, the burden of Data Collection may cause a slowdown in the natural flow of the process.

If such events are affecting your improvement efforts, you need to alter your Data Collection plan.

- The attitudes and perceptions of the data collectors can affect what they see and how they record data. If there is a sense that the data will be used against them, workers may use the data collection process to cast a favourable light on the process being studied. You have to get past this fear in order to collect accurate data. You might want to consider an amnesty programme.
- Data collectors need to be assured that their leaders realise that the data gathered in the past may have been tainted by fear. This requires a commitment by your leadership that the new information—possibly less glowing or flattering—will not be compared against old data or their perception of how your process operated in the past.
- Failure to follow the established Data Collection procedures can add errors to the data. This bias occurs when the Data Collection instructions, training, or checksheets are not adequately prepared and tested in an operational environment. You need to conduct initial training on Data Collection and then perform a small-scale Data Collection trial to see if it all works smoothly. The small-scale trial may uncover some problems which need to be ironed out before you can actively pursue a larger scale Data Collection effort. The trial may reveal that you need to make a minor change in the checksheet to make it clearer or easier to use, or that additional training on Operational Definitions is required to calibrate the eyes of the data collectors.
- Data may be missing. Don't assume that missing data will show the same results as the data you collected. The fact that the data are missing is a clue that they may be different from the rest. It is best to number the checksheets sequentially to make it easier to verify that you have them all and that all the required samples have been taken.

WE USE TO COLLECT DATA

Data are frequently collected using checksheets—structured forms that enable you to collect and organize data systematically. Because each checksheet is used for collecting and recording data unique to a specific process, it can be constructed in whatever shape, size, and format are appropriate for the Data Collection task at hand. Checksheets have three important uses:

- Record information on the key quality characteristics of your process for analysis using tools such as a Pareto Charts, Histograms, and Run Charts.
- Provide a historical record of the process over time.
- Introduce Data Collection methods to workers and supervisors who may not be familiar or comfortable with collecting data as a regular part of their jobs.

TYPES OF CHECKSHEETS ARE THERE

The most common types of checksheets collect data either in tabular form or in a location-style format. Occasionally you may encounter a graphic-style checksheet. No matter which type you are using, make sure that it is clear, complete, and userfriendly.

The three types of checksheets are described below:

1. *Tabular format*: A tabular checksheet—also known as a "tally sheet"—is easy for you and your team to use when you simply want to count how often something happens or to record a measurement. Depending on the type of data required, the data collector simply makes a checkmark in a column to indicate the presence of a characteristic, or records a measurement, such as temperature in degrees centigrade, weight in pounds, diameter in inches, or time in seconds.
2. *Location format*: A location checksheet allows you to mark a diagram showing the exact physical location of a defect or characteristic. An insurance adjuster's pictorial claim form detailing your latest bumper bruise is an example of a location checksheet.
3. *Graphic format*: Another way of collecting data is by using a graphic form of checksheet. It is specifically designed so that the data can be recorded and displayed at the same time. Using this checksheet format, you can record raw data by plotting them directly onto a graph-like chart. An example of a checksheet which also produces a Run Chart as the individual data points are plotted and the adjacent points are joined with a straight line.

WE DEVELOP USEFUL CHECKSHEETS

There is no standardized format that you can apply to all checksheets. Instead, each checksheet is a form tailored to collect the information needed to improve a specific process.

Remember, a well-designed checksheet is the launching pad for an effective analysis in which data become meaningful information.

With that in mind, here are some guidelines to help you develop useful forms:

- Involve the process workers in developing the checksheet for their process.
- Label all columns clearly. Organize your form so that the data are recorded in the sequence seen by a person viewing the process. This reduces the possibility of data being recorded in the wrong column or not being recorded.
- Make the form user-friendly. Make sure the checksheet can be easily understood and used by all of the workers who are recording data.
  - Include brief instructions on the back of the form.
  - Create a format that gives you the most information with the least amount of effort. For example, design your checksheet so that data can be recorded using only a check mark, slant mark, number, or letter.

- Provide enough space for the collectors to record all of the data.
- Designate a place for recording the date and time the data were collected. These elements are required when the data are used with Run Charts or other tools which require the date and time of each observation.
- Provide a place to enter the name of the individual collecting the data.
- Allow enough space so data collectors can write in comments on unusual events. This information could be entered on the back of the form.

WE REVIEW WHAT WE'VE LEARNED

Now let's work through some examples that illustrate how to determine where to collect baseline data and how to use checksheets to capture them.

Example 1: Pharmacy Waiting Time: A team of hospital corpsmen working in a Naval Medical Clinic is attempting to improve service in the pharmacy by decreasing the time patients wait for their prescriptions (*i.e.*, the key quality characteristic).

The desired end state is reduced waiting time for the patients.

The team proceeds as follows:

- They develop a Flowchart of the process. After examining the Flowchart and discussing the steps, they conclude that a possible bottleneck in their process occurs at the point where a prescription is filled by the pharmacist. This is supported by general observations of team members.
- They develop a hypothesis along with their "should be" Flowchart.
- They opt to take baseline data using a checksheet to assess both the time required for the pharmacist's actions and the overall time required to process the prescription.

Example 2: Communications Gateway Site Setup Time: A team of Marines is investigating methods for improving the process to set up a communications gateway site for a tactical data link interface with Navy ships and aircraft. They want to reduce the time it takes to set up the site.

- They develop a Flowchart of their process and review it, identifying possible bottlenecks and slow operations.
- They hypothesize that they can improve their set-up time by focusing on the steps of the process dealing with unpacking the equipment.
- They opt to use a checksheet to take baseline data on that portion of the process as well as on the overall time to complete the installation and checkout.

These two examples illustrate the point that you should collect data from the key process step or steps where the outcome can be most affected.

But beware of making assumptions based only on a Flowchart. You should collect some initial data on all critical steps in the process. Then you should collect detailed data about steps that initially showed great variability, or those that took the longest to perform. Only when these data have been analysed can the cause of pain in the process be identified and acted upon.

SOME EXAMPLES OF CHECKSHEETS THAT WORK

Remember, there are no standard Data Collection forms. Every checksheet is unique to the process you are investigating. Through 22 at the end of this module are examples of checksheets that are well-designed for their purposes. Careful design of your checksheet may allow you to collect two or more factors on a single form.

6

Research Design and Rationale

In this part, which embeds my methodology in the literature and discourse surrounding education research, I explain the issues I explored and the assumptions and rationale that guided my planning and investigations.

RESEARCH METHODOLOGIES

Education research is the systematic investigation of educational phenomena:

- A mode of thinking rather than a shortcut to answers.

Historically a spectrum of methodologies have been applied to education research. These include scientific experimentation, curriculum development, action research and ethnographic inquiries. At one extreme of this spectrum are methodologies based on what Nisbet terms the "agricultural model"— controlled experiments to improve the product. For a long time education research was dominated by this model which aspired to scientific precision in the behavioural sciences.

Kerlinger argues that science aims to "establish laws, systematic explanations or relations that apply generally". This paradigm is based on a positivist view of knowledge–that the truth exists and can be uncovered empirically by research. Lincoln and Guba suggest that this paradigm belongs to a "positivist era" of faith in science and the scientific method. Although we are now in a "postpositivist era", positivism has been "remarkably pervasive". Over the last two decades, however, there has been growing disenchantment in education research with methodologies rooted in this paradigm.

One concern is that the very nature of the experimental method with its laboratory conditions and controls renders the educational situation artificial.

This leaves the relevance of the results to educational practice open to challenge. Another issue concerns the uses and interpretations of inferential statistics, which underlie the experimental method in education research. For example, in recent editorials of the *Mathematics Education Research Journal*, Ellerton initiated debate and invited further discussion on statistical significance tests in education research. In this ongoing debate, the relevance of these statistical tests to the field of education is questioned. Finally, the problems of replicability and generalisability are considerably more complex for human situations than for agricultural products. The context of any educational phenomenon is unique and multi-faceted.

For these and other reasons, there has been considerable movement in education research towards qualitative methodologies that involve open-ended inquiry–based on anthropological models or sociological models.

This paradigm is loosely described by Nisbet as:

- Go and live there and see what it is like.

Qualitative methodologies have interpretation and exploration as the rationale for data collection and analysis.

Some major characteristics of qualitative methodology are described by Merriam as follows:

- The primary instrument is the researcher;
- The research involves natural settings or fieldwork;
- The results are qualitative description, that is "thick" rich text which gives depth to the experience;
- The inductive rather than deductive method is followed, that is abstractions are built from examples.

The methods by which qualitative research is expressed are diverse and may include interviewing, participant observation and content analysis as well as statistical analyses.

To Kirk and Miller:

- Qualitative research is an empirical, socially located phenomenon, defined by its own history, not simply a residual grab-bag comprising all things that are 'not quantitative'.

To me the essence of the qualitative approach to research is the contextualising of the data.

As Van Maanen explains:

- The data developed by qualitative methods originate when a researcher figuratively puts brackets around a temporal and spatial domain of the social world.

This awareness of context implicitly signifies that the research framework is non-positivistic–the research cannot uncover laws about humans which apply at all times and places. This is particularly important for research inspired by Vygotsky's insights where human processes are "historically and socially determined".

In my investigation, the context in which students learn Statistics is a focus of my analytic understanding of their learning processes. Gill points out the high degree of contestation over different research methods and research methodologies. In particular there is rivalry between qualitative and quantitative approaches to research.

Rather than dialogue or debate, Gill likens the rivalry to staking a claim in goldmining days:

- Stake your claim of new ground and defend it by deriding all previous efforts.

In particular, Gill finds the objection to quantitative approaches, troubling and usually irrelevant to the researcher's actual purpose of justifying an appropriate qualitative method.

I agree with Gill that what is significant in any research methodology is the relationship between the findings and the theorising, rather than what research methods were used. If the assumptions underlying the data collection and analysis are poorly articulated then this relationship may not be well developed.

For example, misunderstanding of assumptions underlying statistical analysis can lead to a reliance on numerical outcomes and the reporting of such numbers as ends in themselves. Much of the debate about the role of statistical tests of significance, centres around the misapplication of statistical tests of significance. For example, these are sometimes misinterpreted as validating or invalidating the truth of a research hypothesis. However, poor research is no more inherent in quantitative methods than in qualitative methods. An uncritical acceptance of "highly individualised" narratives can be just as implicated in obscuring the situation of the research in its broader social context as an uncritical acceptance of numbers.

To Gill the adoption of a mixed methods approach is often more productive than a single line of inquiry.

Further, Gill suggests that:

- The schism in education research is a construction–and a construction that is fairly loose and somewhat insecure.

My position agrees with this. To me, different research methodologies have in common a search for meaning and understanding. The various research approaches are by no means mutually exclusive. Many education studies today use both controlled experimentation and qualitative exploration. Moreover, I believe that methodologies applying qualitative or quantitative methods are not even sharply separable. For example, the laws developed using quantitative methods are always statistical, that is concerned with variation–probabilistic. They are not absolute statements but conditional and uncertain. This is not only true of behavioural sciences, but also of the so called "exact" sciences, such as physics.

Precision and objectivity are always a matter of degree. "Fuzziness" is inherent in all processes, be they movements of an atom or a student's performance in an examination. Further, experimental results cannot be divorced from the researcher's carrying out the observations, nor from the context in which he or she does so. In short, all analysis involves interpretation. Mathematical statements may seem impersonal and indisputable but their interpretation is embedded in the rich soil of human experience. Moreover, while, positivist "laws" hypothesise and test relations at the collective and abstract level, they are useful in assisting understanding of the profile and social experience of the individual. The metaphor of a spectrum of methodologies conveys the idea of shedding light. That, in the end, is surely the aim of any education research.

MY METHODOLOGY

My primary aim is holistic and descriptive interpretation, aiming for depth of understanding. Consistent with a Vygotskian perspective, I consider the sociocultural context to be inseparable from individual actions. Hence my primary concern is to interpret the network of relationships between learner, subject matter and context. In attempting to make the data meaningful, I have drawn on various techniques of data collection and analysis. I have found it useful to combine qualitative and quantitative methods.

Overall, my mode of doing research is consistent with what Lincoln and Guba term the "naturalistic paradigm" in that I take as axiomatic that: realities are multiple; myself as researcher and my objects of research are inseparable; my working hypotheses are bound by time and context and all entities are in a state of mutual simultaneous shaping. Moreover, my inquiry is "value bound". That is, my investigation is bound by my assumptions, theories and perspectives and is regulated by both cultural norms and my individual beliefs. For example, my focus on the desirability of deep, rather than surface, approaches to learning and on personal development as part of the educational process reflects my commitment to these values.

The methodology I have used in this project is best described as case study research. Case study research does not claim any particular techniques for data collection or analysis. It focuses on a clearly defined area of investigation– what Smith calls an "instance of a more general class of events".

The case study is appropriate when "interpretation in context" is essential to understanding the phenomenon under investigation. Yin defines the case study to be:

- An empirical inquiry in which the number of variables exceeds the number of data points.

This means that the phenomenon under investigation is complex and bound up with the context, so that the important and interacting variables are too composite and too numerous to isolate and measure. Miles and Huberman sometimes prefer the word "site" to "case" to emphasise that case studies take place in a specified social and physical settings. My design emphasises insight, discovery, description and interpretation within a bounded context.

As with any style of research, reliability, validity and generalisability are prime concerns. However these terms have different meanings according to the position of the researcher. In a positivist framework there is an assumption that there are "true" scores which, given a good enough instrument, can be measured reliably and with validity. That is, the measuring instrument should capture accurately and with stability the true essence of what is being measured.

From a non-positivist perspective, however, there are no absolute truths or universal laws about human processes which can be captured. In such frameworks, the ways in which research findings are validated are socially construed. Constructs represent a consensus by the group concerned. This perspective suggests to me that the criteria for reliability and validity in education research depend on the recognition of the community of education researchers. These criteria are constituted by a convergence of social opinion about what is being investigated and how well this is done.

My understanding of social consensus as the basis for the research criteria accords with the qualitative approach to education research. In particular, Merriam formulates questions about reliability, validity and generalisability as they relate to qualitative case study research. These questions suggest criteria which fit with my framework. In what follows I describe these criteria and give an overview of my attempts to satisfy them. The details of how these attempts were carried out will be described in the context of Study One and Study Two.

Merriam formulates the question of reliability for qualitative case study research as being:

- Are the results consistent?

I have tried to provide a basis for recognition of consistency by means of different modes of "triangulation". This surveying term, now commonly used in education research, was first appropriated by Webb et al who spoke of subjecting findings "to the onslaught of a series of imperfect measures". This could suggest, however, that the perfect measure exists.

In contrast, from my perspective, triangulation is about different ways of viewing a phenomenon, each of which is embedded in its own cultural and social context. Together, these ways provide a multi-faceted perspective on the phenomenon under investigation and better opportunities for understanding it than one method alone.

My modes of triangulation are listed below:

- I used triangulation "between methods", that is different methods of data exploration. Both qualitative analyses and quantitative explorations were utilised to find congruencies and anomalies;
- I also applied triangulation "within methods". For example, interview data was used to supplement, qualitatively, students' written responses to open-ended survey questions and I used a range of statistical analyses to explore patterns in the data;
- My third mode of triangulation was triangulation "between investigators". I checked findings with fellow researchers, informally in discussions and formally by independent classifications and peer reviews.

The prime question of validity for case study research suggested by Merriam is:

- How credible are my findings given the data I have? Can others recognise that my conclusions are supported by the data?

I have tried to achieve internal validity by:

- Making overt my assumptions and procedures and laying an “audit trail” so making it possible for other researchers to follow how I got from the data to my findings;
- Making explicit the context, my personal biases and the theoretical framework;
- Suggesting and checking alternative explanations for my findings;
- Most importantly, submitting preliminary analyses for publication, or presenting my findings at conferences, thereby ensuring peer examination of my research.

In order to satisfy user or reader generalisability that is to enable researchers to apply the findings to their own research, my aim is to make the data accessible– to create a virtual reality so that others can experience the phenomenon through my eyes. The passport to generalisability of my research is its “shock of recognition” by others.

I have tried to achieve this by:

- “Thick” description facilitated by my personal and deep relationships with some of the participants;
- Diverse sites for comparison, namely the Mathematics Learning Centre the Psychology II class and a General Statistical Methods class;
- Inclusion of both unusual and “typical” members of the group being studied;
- Sampling to capture diversity within the project;
- Submersion in the research situation;
- Making explicit the “transferable *theory*” from the investigation.

GROUNDED THEORY

Beginning with Glaser and Strauss there has been much discussion of “grounded research”, that is, the need for the research framework to evolve rather than to be imposed. In my case, a framework of beliefs, shaped by my teaching experience, previous research and interest in activity theory, resulted in my developing a loose theoretical framework early in the project. This was revised and changed at every stage of the research. Study One was explorative based on observations and intuitions gained by many years of teaching experience, in particular, teaching the Psychology students who attended the Mathematics Learning Centre for assistance in Statistics.

The framework developed in Study One then served as the platform for developing the questions, hypotheses and procedures for the broader Study Two. This conceptual development is ongoing, informed by discussions and the act of writing this thesis and submitting research papers for review. Throughout the investigation, I have attempted to give a voice to the students. My analysis, though influenced by my experiences and by my interpretation of activity theory, is grounded in the perceptions expressed by the students.

ANALYSIS OF STUDENTS' APPRAISALS OF LEARNING STATISTICS

This part outlines the method used to analyse the first open-ended survey question:

- Would you study statistics if it were not a requirement of your psychology course? Please give reasons for your answer.

As I specified the first part of the question was simply analysed in terms of whether the student answered "Yes" or "No".

The reasons that students gave for their choice were categorised in the following way:

- I started with the categories that had been identified in my pilot study. However, I expected that differences would emerge in Study Two. The pilot study had been carried out three years earlier. A different lecturer taught Statistics then and the pilot study was conducted at the beginning of the academic year, so that students had less experience of Statistics than the participants in Study Two. Moreover my pilot study sampled not only Psychology II students but also students of General Statistical Methods. Nevertheless, many of the same categories were found to be useful for Study Two.
- Most students gave several reasons for their decisions. For example, some students who reported that mathematics was boring also believed that it was hard. I classified each portion of the student's response into one of the specified categories. If the student gave a reason which indicated a negative view, I assigned her a score of –1 for the appropriate category. I assigned a score of 1 for favourable observations. If the student made no response in a particular category, a score of zero for that category was allocated. In this way students could have non zero scores in a number of categories.
- Each student scored either –1, 0 or 1 for each of the categories. Hence the total number of favourable responses and unfavourable responses in any one category was equal to the number of students whose reported reasons were classified in that category.
- If the student's response did not fit into any of the categories identified so far, a new category was specified.
- When all 279 surveys had been classified, some of the original categories were found to be empty and dropped.

ANALYSIS OF STUDENTS' CONCEPTIONS OF STATISTICS

I based my analysis of students' conceptions of Statistics on the modifications of the phenomenographic method that were developed in earlier research by my colleagues and myself. As was the case in those investigations, the categories of description and their patterns of distribution are considered to be major results of Study Two. The phenomenographic analysis of the data on students' conceptions involved three stages. In the first stage the categories of conception were identified.

The second stage consisted of classifying the 279 surveys into these categories. In this process the categories were clarified and refined. In the third and later stage I checked each survey response to pick out any that I felt had been misclassified and submitted these to further examination and discussion.

Stage One: Identifying the Categories of Conception

The first stage in the analysis of the data was to identify a set of qualitatively different categories of description of students' conceptions of Statistics.

This involved the following procedure:

- An initial set of categories was identified by myself and another education researcher, Mr Peter Fletcher, experienced in phenomenographic research. This was achieved by our independently reading and classifying the entire set of 279 written responses to the open-ended question below:
  - What in your opinion is this statistics course about? Please explain as fully as possible.
- We two researchers then compared and discussed our initial categories and agreed on a draft set of categories and sub-categories.
- Together with a third researcher, we independently classified 30 selected responses in terms of this draft set of categories.
- The three of us compared and discussed the individual classifications for the 30 responses and agreed on a final set of clear statements for each category.

The categories were arrived at by negotiation and discussion. We looked at commonalities in the students' responses and differences between them.

Stage Two: Classifying the Surveys

All 279 responses were then classified into the identified categories by myself and my co-researcher, working independently. This classification was done in the following way.

- For each student the response to Question 3 was read in isolation to the rest of her survey.
  - What in your opinion is this statistics course about? Please explain as fully as possible.

  An initial decision was made as to the category the response best fitted. I was mindful of the fact that, although I had phrased the question in order to elicit the student's own perception of the subject matter, some students were likely to interpret the question as asking about their perceptions of the teacher's view of Statistics, rather than their own awareness.
- Each of us re-appraised our initial decisions in the light of students' responses to the other open-ended questions, namely:
  - Question 1 Would you study statistics if it were not a requirement of your psychology course? Please give reasons for your answer.
  - Question 2 Think about the statistics you've done so far this year.
    a. How do you go about learning it?
    b. What are you trying to achieve?

In this way we amplified our understanding of each student's awareness and perceptions of Statistics and gained further insight into her way of experiencing the knowledge. We individually recorded a final category for each student's conception of Statistics by taking account of the student's overall responses to all three open-ended questions. The initial category that had been identified from Question 3 alone played no further part in the analysis, but was found to be very useful in the ensuing discussions between myself and fellow researchers in making explicit the reasons for the final categorisation.

- We then discussed all classifications and reached agreement on any that did not match. In this process of classification and discussion the categories themselves were further clarified.

That is, as Marton explains, the analysis:

- Is dialectical in the sense that meanings are developed in the process of bringing quotes together and comparing them.

Not only core meanings are important in specifying the categories, but, as Marton points out, "borderline cases" play an important part in delineating the categories. In accord with Marton's phenomenographic method, the categories transcended the boundaries between students' responses to different questions and also transcended the boundaries between individuals. That is, the students' separate responses to different questions and their individual ways of expressing themselves were subsumed into general categories.

Stage Three: Triangulation and Checking

Triangulation on the analysis of students' conceptions of Statistics was achieved by:

- Interpersonal agreement between at least two researchers;
- Interviewing selected subjects;
- Peer review. I presented my ongoing analysis at three research conferences, conducted approximately six months apart. In this way I obtained feedback on my ideas, methodology and findings as my study progressed.

After a period of some months I read through all the responses, one category group at a time. On the basis of the better understanding of the data now available to me, for example through interviews, I discovered a few responses that appeared to me to have been misinterpreted in earlier discussions. These were further discussed with my colleague and reclassified in some cases.

DEVELOPMENT OF THE APPROACHES TO LEARNING STATISTICS QUESTIONNAIRE

The Approaches to Learning Mathematics Questionnaire and the Approaches to Learning Statistics Questionnaire were derived from the two scales of the Study Process Questionnaire which denote surface and deep approaches to learning. While the Biggs' scales purport to investigate approaches to studying in general, the Approaches to Learning Mathematics Questionnaire investigates students' approaches to learning mathematics, as a generic field of study. That is, in the Approaches to Learning Mathematics Questionnaire, the wording of Biggs' deep and surface scales was changed to refer to approaches to learning mathematics, rather than to studying in general. My Learning Statistics Questionnaire from which the items of the ALSQ were selected, was modified from of our Approaches to Learning Mathematics Questionnaire by changing "mathematics" to "statistics" in the items. The Learning Statistics Questionnaire refers to students' approaches to learning Statistics–as a particular subject studied for Psychology II. Hence the three questionnaires: the Study Process Questionnaire; the Approaches to Learning Mathematics Questionnaire and the ALSQ represent a progression–each focusing and particularising the constructs of interest in previous investigations.

The Learning Statistics Questionnaire completed by the Psychology II students contains 28 items on approaches to learning Statistics. The 14 odd numbered items were derived from items in the previous questionnaires indicating a surface approach to learning and the 14 even numbered items were derived from the scales denoting deep approaches to learning. For each item a choice of responses numbered 1 to 5 was provided. The lower end indicates that the student "only rarely" adopts this approach when studying Statistics, while the upper end indicates that the student "almost always" does so. Hence, a student's score on each subscale signifies how usual it was for her to adopt the specific approach to learning Statistics.

I carried out item factor analyses and scale reliability analyses in order to determine the structure and internal consistency of the items. As a result of these analyses the ALSQ was defined as consisting of 18 items, twelve making up the Deep Scale and 6 items constituting the Surface Scale. Scale validity of the ALSQ is indicated by relationships with other variables such as willingness to learn Statistics.

ETHICS

This study was approved by the University of Sydney Human Research Ethics Committee. In order to ensure that I fulfilled my ethical obligations to the students and academics participating in my research I addressed the following issues.

Voluntary Participation

Survey

In the case of the survey, I invited students to participate both verbally and in writing. I explained the purpose of the research and how their responses would be used. Students who did not wish to complete a portion of the survey, or indeed the whole survey, were at liberty to desist. A few students omitted to give their Student Identity Number and one person refused to complete the survey. No pressure was brought to bear on them.

Interviews

Only students who gave their written consent to be interviewed were contacted by telephone. All interviews took place on a face to face basis and were audio taped. Hence the participants were able to see when the tape recorder was running, and I invited them to switch the recorder off at any time if they so wished.

I asked students to sign a consent form. I also signed this form assuring the student of confidentiality and explaining again the purpose of the research. During the interview each participant was given the opportunity to ask any questions and discuss any difficulties relating to Statistics. Some students took this opportunity to ask questions relating to the course content, the assessments or my research. Students were offered full editorial rights over the transcript of their own interviews. None opted to take advantage of this offer.

Disclosure of my Aims

I informed all participants of my aims: in writing, on page one of the survey, on the interview consent form, and verbally. Teaching staff were invited to discuss the research with me at various stages of preparation for the project and analysis.

Guarantee of Confidentiality

All students and staff participating in my research were guaranteed confidentiality. False names are used throughout. As recognition of students by staff members close to them is impossible to prevent, no excerpts from students' interviews were published during the year the students were in Psychology II.

Opportunity to Complain of any Ethical Misconduct

Students were given written information as to whom any complaints or reservations about any aspect of their participation in this research could be addressed.

Taking Account of Power Structures

As a staff member of the Mathematics Learning Centre, I have no input whatsoever over students' grading and am not involved with tests and examinations in any way. Similarly I have no authority over other academics. My job is to assist students at their own behest with any subject at The University of Sydney which involves elementary level mathematics and statistics. Students and teachers were invited to assist me by participating in this research.

RATIONALE FOR STUDY ONE

A focus of my inquiry is to interpret and extend the theory of activity by describing the learning of Statistics as purposeful and interactive activity that develops in a sociocultural context. As is the case for most students of the Mathematics Learning Centre, these students were at least 25 years old and had been out of the education system for a number of years.

The rising number of older students in many universities means that research on their learning is becoming increasingly relevant and significant. In addition, observations made by mature adult students are insightful and supply valuable material for stimulating reflections on teaching. The five students who were invited to participate in Study One were selected purposefully. As will be explained in these students had diverse life experiences, a range of academic backgrounds and singular personal attributes which led me to believe that their participation would be valuable to my research. In short, I believe that an understanding of the perceptions, actions and evaluations of the participants of Study One is helpful in shedding light on important issues concerning students who study statistics at university as a service course, rather than by choice. In keeping with a qualitative research design, my analysis of this data is interpretative rather than explanatory, acknowledging the complexity of the issues faced and aiming to indicate the dimensions of that complexity, as well as to stimulate reflection and dialogue.

MODEL FOR METHODOLOGY IN STUDY ONE

The design of this study reflects my consideration that the naturalistic paradigm has the best fit to my aims described. That is, in this study, with one exception, I have adopted the attributes of the naturalistic paradigm. The exception concerns my lack of adherence to grounded theory. What I studied was, to some extent, delineated in advance, informed by my understanding of activity theory.

The attributes of the naturalistic paradigm are detailed with a description of how I complied with each in Study One.

- *Characteristic* 1: Natural Setting
  - Study One was carried out in the Mathematics Learning Centre in the context of my teaching the five students during their second year of study of Psychology. All the students were regular and intensive users of the Centre and in some cases spent five or more hours per week working with me on their Statistics. Indeed the students frequently expressed the idea that they “lived in here” during their second year of study of Psychology. My observations of them for the purpose of the study are inseparable from my actions designed to assist them with their learning.
- *Characteristic* 2: Human Instrument
  - I consider myself to be the primary data gathering instrument. That is, the evaluation of what the students said, did or wrote, is personal and subject to my own perceptions, beliefs and values.
- *Characteristic* 3: Utilisation of Intuitive Knowledge
  - The study reflects my deep and warm relationship with the students. My analysis takes account of my experience of teaching them and others who have sought my assistance in learning Statistics.

- *Characteristic* 4: Qualitative Methods
  - Qualitative methods of data collection and analysis were used, namely descriptive and exploratory techniques aiming to illuminate multiple truths in context.
- *Characteristic* 5: Purposive Sampling
  - The students were not selected as being representative of any general group of students. Rather, it appeared to me that due to their singular characteristics exploring these students' orientations to learning Statistics would be of value in understanding how students learn Statistics.
- *Characteristic 6:* Inductive Data Analysis
  - I have tried to construct general themes from the particular examples provided by the students' accounts.
- *Characteristic* 7: Grounded Theory
  - I did not wholly comply with the description by Lincoln and Guba of having "the substantive theory emerge from the data", as I was building on activity theory. However, my theory evolved as the study unfolded.
- *Characteristic* 8: Emergent Design
  - The design unfolded as I observed the students. It was influenced by my interactions with the students. My research questions evolved from informal observations and intuitions. They were incorporated into the questionnaires completed by the students during their study of Statistics and thence into the semi-structured interviews conducted with the students after they had completed second year Psychology. Themes emerged in these interviews and subsequent interactions with colleagues and experts in education. The design of Study One was only fully completed when I had written and revised a paper on the study for publication.
- *Characteristic* 9: Negotiated Outcomes
  - My interpretations were discussed with the five participants during their interviews and re-evaluated together with peers and other researchers. I also tested my working hypotheses in other studies. I did not, however, ask the five participants to evaluate my interpretations of their interviews. I would have preferred to do so, but there were practical and ethical problems. By the time they were interviewed the students had already completed second year Psychology, and so had ended their association with me at the Mathematics Learning Centre. Not all of them were accessible to me afterwards. Also, I did not wish to exploit my personal relationship with the students by seeking their cooperation in my project beyond our negotiated agreement.
- *Characteristic 10:* Case study reporting mode
  - I adopted this mode of reporting, also called the "narrative format" in order to constitute the stories of the students, as I perceived them.

In this way of reporting I was influenced by Bruner's narrative mode of thought, as explained. This genre recognises the importance of description and context and is intended to make possible understanding on the part of the reader.

- *Characteristic* 11: Idiographic Interpretation
  - I make no claims that the study fulfils any general human condition. The data are interpreted in a particular situation and this interpretation is dependent on my personal perspective of the context.
- *Characteristic* 12: Tentative Application
  - While I hope that my research will be of value to other education researchers and to practitioners of teaching statistics, my findings are confined by temporal and physical boundaries - a snapshot of a situation. However, I have taken what steps I can to ensure that the extent of transferability and applicability by others can be judged by others.
- *Characteristic* 13: Focus-Determined Boundaries
  - The boundaries of this study are determined on the basis of the theoretical concepts which frame the study, and by the particular activities undertaken by the students in the social setting.
- *Characteristic* 14:Special Criteria for trustworthiness
  - I have tried to meet the criteria of reliability, validity and generalisability. The aim of meeting these criteria is to enable my conclusions to be recognised and used by others.

PREPARATORY RESEARCH FOR STUDY ONE

Two preliminary studies informed Study One. The first investigation formed the basis of the initial data collection on four of the participants in Study One. This investigation addressed students' attitudes to and beliefs about mathematics and statistics. The students were surveyed during a bridging course in statistics.

The course was advertised as a preparation for those students lacking the basic mathematical skills needed for their study of statistics in second year Psychology or for postgraduate courses in Public Health. The participants in the bridging course were self selecting: each of the 60 students enrolled felt poorly prepared for the statistics course which was a compulsory component of his or her chosen course. One interesting finding of this preliminary study related to differences in students' perceptions of statistics as compared to mathematics. In response to free association with the words "statistics" and "mathematics", students reported that statistics is "useful" and statistics is"misleading". Disraeli's saying: "Lies, damned lies and statistics!" evidently rings true for many people. Mathematics, on the other hand, was seen by these students as being either true or false. I took these perceptions into consideration in my analyses of Study One and of Study Two. That is, I paid attention to differences in the students' usage of the words"statistics" and "mathematics" in their written responses to open-ended questions and in interview data. The second investigation looked at the attributes and perceptions of Psychology II students who attended the Mathematics Learning Centre: their motivation for studying Statistics, their reasons for using the Centre, their background in mathematics and their evaluations of the learning environment. Correlations between the students' examination grades and their ages, prior levels of mathematics studied and performances in other areas of Psychology were addressed in this study.

I discovered relevant questions to be addressed, for example: What were students' expectations about themselves as learners of Statistics? I also identified methodological issues from this investigation. For instance, I applied the "action research spiral" to this investigation. At the same time as building up the pragmatic and methodological foundations for Study One, I was building up a theoretical framework. This was based on the ideas of Vygotsky and Leont'ev as I have related. I applied their theories to the context of the Mathematics Learning Centre by illustrating the major features of the theory of activity with examples taken from my teaching practice.

I also developed a model for teaching these students based on my own, adult, experiences of learning to ice skate. In order to view the activity of learning Statistics from the students' perspectives, I related my own deficiencies in the three C's needed for ice skating, to the students' feelings of inadequacies about themselves as learners of Statistics. I compared my ways of operating in the zone of proximal development with the support of the skating instructor, to the ways students learned to overcome their difficulties with my support. In these ways–by developing theoretical concepts, intuitive knowledge and metaphors for my teaching and learning, and submitting publications on these to peer review, I developed my themes and working hypotheses for Study One.

METHOD FOR STUDY ONE

Participants

Each of the five participants of Study One had exceptional attributes in some respect. The names I use to refer to the students, are fictitious but indicate gender.

Norman, Sandra and Alice were extremely successful students of second year Psychology:

- Norman and Sandra both achieved High Distinction grades, a rank attained by only four per cent of their Psychology II class of 473 students, while Alice gained a Distinction for the course. Alice and Hettie had the unusual characteristic of not having studied mathematics at secondary school level at all. In fact Hettie's school education ended six months into secondary school, making singular her later academic achievements. Both succeeded in passing Statistics after a considerable struggle although both had high grades in other areas of the course. Four of these students were pursuing a study of psychology for vocational reasons. Ernest, on the other hand, was of an age where vocational interests are not relevant and was studying psychology to satisfy his intellectual curiosity. This resulted in his having a different attitude to his studies.

Data Collection

My data collection for Study One included informal observations made while participating in the students' learning, audio taped interviews, short surveys and questionnaires relating to the students' attitudes to and strategies for learning Statistics, the students' written evaluations of the teaching and environment of the Mathematics Learning Centre, demographic information and assessment results. 1As is customary with qualitative research my data consists of descriptions, direct quotations and excerpts, obtained by close psychological contact with those being studied. The major sources of these data were interviews conducted after the students had completed the course. The students trusted me and were uninhibited in the interview situation. Some of the students' quotations which I report were obtained from questionnaires completed by them at various stages of the academic year and from their written notebooks.

My initial data collection consisted of survey responses of students taking a bridging course in statistics. Four of the participants in Study One took part in this bridging course. My aim was to elicit the students' own understandings and perspectives instead of imposing my own ideas. Accordingly, at the start of the first session, without more than a brief personal introduction, the students were asked to respond freely with their associations to "mathematics" and "statistics".

These open questions were followed immediately by a more structured questionnaire. From this investigation, I obtained an initial idea of the issues pertinent to the four students' views on mathematics and statistics. Demographic information about the five participants and their reasons for seeking help in Statistics were obtained from their enrolment forms for the Mathematics Learning Centre. This was completed by the five students early in the first semester.

Throughout the year I observed the students informally, while I was engaged in teaching them individually, or while they were working in small groups in the Centre. As each of the students spent considerable and regular time with me according to their own needs, I had many opportunities to view their written work and to observe their strategies for learning. I spent considerable time interacting with them, both informally and in formal instruction. A survey consisting of an open-ended questionnaire and Likert-type items was completed by the five students in semester two. Information on their evaluations of the Mathematics Learning Centre, as the setting for learning Statistics, was obtained from Mathematics Learning Centre Evaluation forms for the students, all of whom opted to sign their names on this survey.

After their final examinations in second year, at the conclusion of the academic year, I invited each of the students to come into the Centre for a long interview. According to Patton the purpose of the interview is to enter into the other person's perspective. To be successful, the interviewer must be neutral, that is non judgemental, and have rapport with the subject. The protection of the respondent must be ensured and ethical issues taken into account. The logistics of the interview must be worked out. I based my interview questions on the activity theory framework and on my data analysis of the students' responses undertaken so far. The interview questions were designed to gain an in-depth knowledge of the students' goals, perceptions, attitudes and beliefs, relating to the Statistics course which they had recently completed. The interviews were semi-structured but neither the exact wording nor the order of the questions was precisely determined in advance. I allowed for feedback, interactive reassessment and probing.

Furthermore, I took advantage of the students' relaxed frames of mind and the interpersonal trust we had established, to follow up any unseen issues that arose in the course of the interview. I was also able to explore new openings suggested by the students' responses. In short, I acknowledged the sub texts of the conversations. The interview questions were divided into four broad areas, as advocated by Patton.

These were about:

- *Behaviour or Experience, for Example:* Did any changes in your learning pattern occur during the year?
- *Opinions and Values, Such as:* How useful do you think what you have learned in Statistics will be to you?
- *Emotional Responses and Feelings:* How did you feel about mathematics then ?
- *Background Demography, for Example:* Have you done any mathematics or statistics since school, apart from your psychology course?

The ways in which the questions were framed included four modes, described by Strauss, Schatzman, Bucher, and Sabshin.

These are:

- *The Hypothetical Mode:* For example–If Statistics had not been assessed, but nothing else changed, how would you have approached learning it?
- *The Devil's Advocate Mode:* For example–Some people would say, that if you can't cope with statistics you have no business doing psychology. Would you comment on that?
- *The Mode of Posing the Ideal:* For example–What to you would be the ideal statistics course?
- *The Interpretative Mode:* What do you mean by?/Could you explain?

I used this mode for probing. The interviews were audio taped and I transcribed them myself, as a way of getting close to the data. I also collected and read the students' notebooks in which they had made their own summaries of Statistics. This added further insights into each individual's ways of working and feeling.

For example, Alice prefaced her written summary of a part she found difficult with:

- Analysis of Variance G-d help me!

The notebooks were returned directly and promptly to the students but I retained confidential photocopies of illuminating parts.

Data Analysis

The aim of the data analysis is to tell the story of the five students in a way that could be recognised by other researchers and educators. As advocated by Sieber each stage of the data collection was analysed and suggested further data collections and analysis in an ongoing cycle. In the first level of analysis I arrived at categories of description, namely nodes or clusters of similar objects or characteristics, by reading systematically through the case records and finding patterns. In particular, the act of personally transcribing the audio taped interviews suggested categories of similar phenomena. Then I found units of information: for example, a sentence or notion that stood by itself, and I classified each into a category.

In this way the categories were clarified and refined to ensure that they were, as proposed by Holsti and Merriam, relevant, exhaustive, mutually exclusive, independent and formed from a single classification principle. The categories I arrived at were: demographic information; students' goals with respect to learning Statistics; their perceptions of school mathematics; perceptions of "real life" statistics; perceptions of context in the activity of learning Statistics and approaches taken to learn Statistics. The categories were entered on a database as a precursor to my writing an integrated and interpretative description.

That is, from the earliest stage of analysis, my database was a narrative, intended to be read by an audience of educators. Each stage of the analysis corresponded to a refinement of this narrative–from disjointed elements, through motifs to a unified whole.

The second stage of the analysis was to identify themes which connected the categories. The themes were derived from what the students said–common threads in the interview data. These were analysed in terms of Leont'ev's breakdown of activity into three levels: the milieu or social setting of the activities, the goals directing the actions and the operations or automatic components of the actions. One major theme which emerged, relates to students' evaluations and self regulation of the task of learning Statistics. The next step was to formulate working hypotheses.

These were outlined are reproduced below for the convenience of the reader.

- The students identified two different sorts of statistics or mathematics–school mathematics and "life" statistics or mathematics.
- Actions were goal directed.
- The students' approaches to learning Statistics were related to how they positioned themselves with respect to the learning task.
- The students' metacognitive processes of monitoring and evaluation were paramount in defining their activities.

The last stage of analysis involved integrating my interpretations and explanations into a coherent description. The results of this study served as the starting points for Study Two.

ANALYSIS OF DATA

CRITERIA

Sieber suggests that good analysis usually involves the following:

Intertwining of analysis and data collection; Formulating classes of phenomena–that is, categorisation of concepts; identifying themes–linking concepts, noting regularities, patterns; provisional testing of hypotheses–looking for concomitant variation, ruling out confounding factors, identifying intervening variables. I attempted to meet these criteria in each of the studies.

TRIANGULATION USING QUALITATIVE AND QUANTITATIVE METHODS OF ANALYSIS

Qualitative and quantitative analyses each have assets and liabilities. The aim of triangulation is to exploit the assets of each method while minimising the deficiencies of each. To me, the singular features of quantitative analyses are the level of abstractness at which the findings can be described and the focus on pattern rather than detail. These generate a powerful way of communicating a complex and dynamic system. Quantitative data are useful for discerning distributions, trends, relationships and anomalies. The strengths of qualitative analyses are the very features lost in quantification–the richness and depth of description concerning the particular individual and context.

As Weis points out:

- Qualitative data are apt to be superior to quantitative data in density of information, vividness, and clarity of meaning.

My application of complementary methods of analyses produced opportunities for identifying consistencies within the data, thus enhancing my confidence and encouraging better definitions and enrichment of the description. They also produced surprises and anomalies. These led to alternative and deeper analyses and more insightful interpretation and understanding of the data. Most importantly, the complementary methods allowed for my immersion in the data. This resulted in my improved closeness to the data.

Methods of Triangulation

As a researcher, my task is to gain insights into the data by systematically interpreting their meaning. Different methods of inquiry were used in Study One and in Study Two to investigate the network of relationships relating to the research problem. These include holistic description and statistical explorations. In both studies I have triangulated the analyses "between" and "within" methods by using a variety of approaches both qualitative and quantitative in order to better describe and understand the problems in their contexts.

In Study One and in Study Two, "within methods" triangulation of the qualitative data included using material from interviews, surveys, informal observation and students' written work. Quantitative indicators, such as performance on assessments and demographic data, added another dimension to the descriptive data in both these studies. In Study Two, my triangulation "within methods" also included using a range of statistical analyses such as cluster analysis, factor analysis and correlations. The different ways of exploring the data in this study produced patterns and trends which indicated consistency. For example, the qualitatively different categories of students' conceptions of the subject matter, which emerged from students' responses to open-ended questions, were confirmed as having links to students' approaches to learning Statistics, analysed quantitatively, using scores on a Likert-type questionnaire. Strong relationships were found among the variables, such as students' conceptions of Statistics and their willingness to study it.

Statistically significant trends converged with qualitatively significant differences between the categories. Further triangulation for both studies was provided by "between investigator" analysis and interpretation. Here, discussions with colleagues at each stage of the two studies provided valuable feedback on my research design, interpretations of findings and conclusions. Some independent classifications and checks were also carried out on my categorisation in Study Two.

For instance, students' conceptions of Statistics as reported in their surveys, were categorised by myself and, independently, by another education researcher. Since this researcher is male, younger than me and has a different educational background, we brought very different perspectives to bear on the data. Discussions on the categorisation were held with a third researcher, allowing for further negotiation of different views. In addition, as explained earlier I consider it important to submit my work for publication as this enables me to experience peer reviews of each aspect of my project. A list of my relevant publications appears at the front of this thesis. This includes pilot studies and publications on preliminary findings, both for Study One and for Study Two as I carried them out. I will specify these in the context of each of the two studies.

MODES OF THOUGHT

I used a different genre for expressing my analysis in each of my two studies. In this I was influenced by Bruner's two modes of thought. Bruner calls these the "paradigmatic or logico-scientific mode" and the "narrative mode". The paradigmatic mode of thought is driven by "principled hypotheses" which relate not only to what is observed but also to possible, logically generated worlds.

This mode is the basis of science and is characterised by Bruner as:

- Attempts to fulfil the ideal of a formal, mathematical system of description and explanation. It employs categorization or conceptualization and the operations by which categories are established, instantiated, idealized, and related one to the other to form a system.

This is the mode of thought underlying my attempts to build up an "activity system" by a methodical and logical relation of parts to a whole, guided by empirical discovery and "reasoned" hypotheses. In Study Two my style of reporting is consistent with this mode of thought. However, in Study One and in my overall “story”, I am also concerned with the narrative mode of thought. This mode is the basis of literature.

As an art form it:

- Deals in human or human-like intention and action and the vicissitudes and consequences that mark their course. It strives to put its timeless miracles into the particulars of experience, and to locate the experiences in time and place.

The essence of this mode is the attempt to create gripping drama and believable accounts. It relates events temporally as well as by themes. It provides a basis for understanding individual actions and events as part of a unified “plot”. I have no interest in creating drama but I do have a concern with the unfolding of particular students’ experiences and the insights these provide into human conditions. In short, these two modes are both cultural tools, forms of which I have employed in my research activity.

STATISTICS IN EDUCATION

Statistics in education involves the application of statistical methods and techniques to analyze data related to educational processes, outcomes, and policies. It plays a crucial role in informing decision-making, evaluating educational interventions, and assessing the effectiveness of educational programs. In education, statistics is used to collect, organize, analyze, and interpret data on various aspects such as student performance, teacher effectiveness, school demographics, and resource allocation. By employing statistical techniques such as descriptive statistics, inferential statistics, regression analysis, and multilevel modeling, researchers and policymakers can identify patterns, trends, and relationships within educational data, enabling them to make evidence-based decisions and develop informed policies to improve educational outcomes. Moreover, statistics in education aids in monitoring progress towards educational goals, identifying areas for improvement, and measuring the impact of interventions and reforms. It provides educators and policymakers with valuable insights into the effectiveness of educational practices, the allocation of resources, and the distribution of opportunities, ultimately contributing to the enhancement of educational quality and equity. The book on Statistics in Education provides a comprehensive overview of statistical methods and their applications in educational research, assessment, and policy-making, aiding educators and policymakers in making informed decisions to enhance educational outcomes.

Dr. Rajesh Sharma did an M. Sc. in Mathematics from Maharshi Dayanand Saraswati University, Ajmer (Rajasthan), B.Ed. from University of Kashmir and M.Ed. from IGNOU, New Delhi; Currently, he is working as a Associate Professor, Faculty of Education, Tantia University Sri Ganganagar (Rajasthan). He has published more than ten (10) Research Papers in National Journals. He has participated and presented papers in more than (15) International/ National Seminars. He has more than twenty-two (22) years of teaching experience. Under his guidance, 8 students were awarded Ph.D. in Education.

ACADEMIC
UNIVERSITY PRESS

4378/4-B, Murarilal Street, Ansari Road, Daryaganj, New Delhi-110002
Phone : +91-11-23281685, 41043100, Fax: +91-11-23270680
E-Mail: academicuniversitypress@gmail.com

ISBN-978-93-6284-149-0