Computer-Assisted Learner Group Formation Based on Personality Traits

Dissertationsschrift zur Erlangung des Grades eines
Doktors der Naturwissenschaften
am Fachbereich Informatik
der Universität Hamburg

Vorgelegt von
Rahel Bekele
aus Addis Abeba, Äthiopien

Oktober 2005

Genehmigt vom Fachbereich Informatik der Universität Hamburg auf Antrag von

Prof. Dr. Wolfgang Menzel (Universität Hamburg)
Prof. Dr. Christiane Floyd (Universität Hamburg)
Prof. Dr. Darge Wole (Addis Ababa University)

Hamburg, den 21. Dezember 2005

Prof. Dr. Siegfried Stiehl (Dekan)

Dedicated to

my husband Tesfaye Biru

&

my beloved children
Tibebu, Elshadai, Tehut and Bezawit

ACKNOWLEDGMENT

The list of people who provided inspiration and assistance in this research is probably longer than the paper itself. In particular, I would like to thank the following individuals.

Foremost, I was fortunate to have three challenging and supportive supervisors:

> **Professor Dr. Christiane Floyd**: who has been the cause for my Ph.D. work by way of encouraging me to develop my research work into a Ph.D. study, particularly under the circumstances that I could not stay in Germany for the whole period of the study. I thank you for the excellent work environment you provided me with, during my short stays in Hamburg. Thank you also for your concerns, unreserved comments and advices which made me seriously look into the social dimensions of the research work.
>
> **Professor Dr. Wolfgang Menzel**: whose rich knowledge and ideas upon which this study is based. It is due to your guidance that I have a good insight of the conceptual framework and how to come up with the final work of the research. Every time I visited you in your office, I was filled with hope and encouragement on my research. At the end of every meeting, you replaced my frustration with inspiration. Thank you so much.
>
> **Professor Dr. Darge Wole**: Thank you for sharing your rich knowledge in the field of educational psychology, the many helpful discussions, constructive criticisms and ever-ready guidance and encouragement, and most of all for exposing me to various psychological and social science researches.

Without the love and support of my husband **Tesfaye Biru,** I would be nowhere. Thank you for your patience, invaluable fortitude, the endless discussions and arguments, and for bringing my prose up to the appropriate level of editorial quality. You are my mentor and inspiration.

Dr. Tadewos Korotto: I owe you special thanks for the support and advice I received on the statistical aspect of this study. Every time I thought I was lost, I was rescued by your intelligent insight on statistical concepts and applications.

Professor Dr. Ingrid Wetzel: It is your warm hospitality and spiritual advices which kept me strong during my stays in Hamburg. Thank you for everything you have done, and for your willingness to serve as head of the Ph.D. committee**.**

Professor Mogessie Ashenafi: Thank you for your comments on the first draft of the thesis.

My beloved friends Dr. Abebe Rorissa, Bezakulu Atnafu and Dereje Befekadu: I am indebted to you for your continued support and encouragement throughout the course of this study.

Solomon Teferra, Martha Yifiru, Berhanu Beyene, Aida Bahta and Seble Beyene: Thank you for making me feel comfortable during my stays in Hamburg. I owe you so much for many memorable experiences, for providing me with accommodation and for bearing with my disturbances late at night.

Ato Getachew Jemaneh, Dean of Faculty of Informatics at Addis Ababa University and **my colleagues** at the Department of Information Science: I am grateful for your help at various levels of the research work.

Teachers and students in Yekatit 12 Senior Secondary School: Thank you for your cooperation to make this study a success.

I would also like to thank the **German Academic Exchange Service (DAAD)**, the **Austrian Exchange Service(ÖAD)**, **Dr. Per Ekelund** at Åalesund Univesity College (Norway) and **Addis Ababa University** for their financial support.

Last, but not least, my love and special thanks go to my children **Tibebu, Elshadai, Tehut** and **Bezawit Tesfaye,** for bearing with my absences from home and having to keep quiet during my long hours of work at home. I am very much indebted to **Tigist Gebru** and **Tsehay Gashaw** who replaced me in the role of a mother whenever I was not around.

Rahel Bekele
October, 2005

ABSTRACT

This study addressed two main issues in relation to forming effective heterogeneous learner groups to improve student performance. One was the task of developing a performance prediction model without administering exams and the other was the development of a software tool to form effective heterogeneous groups.

Taking mathematics as the subject of the experimental study, the attributes relevant to compose effective groups were identified based on a survey (involving review of literature and data collection) and discussions with experts in the field. The attributes identified were: Gender, group work attitude, interest for mathematics, achievement motivation, self confidence, shyness, English performance and mathematics performance. Findings from the survey also revealed that, being the contributory factors, the first seven attributes can be used to build the mathematics performance prediction model. Once the attributes were identified, an instrument was developed to collect experimental data. The source of these data were 11^{th} and 12^{th} grade students in one senior secondary school in Addis Ababa, Ethiopia.

Based on the experimental data, a Bayesian performance prediction model was developed where a 70.9% prediction accuracy was first observed. Further experiments and modification of the prediction model increased the level of prediction accuracy to 78.4%.

For the purpose of creating heterogeneous groups, a mathematical model was developed. In particular, applying the concepts of a vector space model, each student was represented in a multi-dimensional space by a vector whose features/components were made up of the values of personality and performance attributes associated with the student. Three algorithms were considered for the purpose of creating the groups based on the mathematical model. The algorithm which generated a reasonably heterogeneous group, was then selected for further experiment. The study has also experimented on the approaches for group composition depending on the availability of student records. One approach (based on a batch-processing algorithm) is used where there is a need to create grouping for a set of students (for instance, students in a class). The other approach (incremental group composition algorithm) does not need the whole data set in advance.

In order to evaluate the software tool, an experiment was conducted on three sections of students in the same high school. These students were first given a pre-group work exam. Students of one section were randomly grouped, students in the second section were made to select their own groups, and students in the third section were grouped by the software developed.

These three sections of students were allowed to study in groups for a period of eight weeks and a post-group work exam was administered. Some of the statistical tests applied were

- the paired samples T-test: to test whether there is a significant difference between the pre- and post- group work exam results;
- a regression analysis: to explain the relation between total hours of group work attendance and change in level of performance;
- test of difference of proportions: to compare between the grouping methods;
- test of difference of two means: to test the stability of the incremental version as compared to the batch processing;

The experimental results confirmed that

- students grouped based on level of performance and personality traits/attributes perform best as compared to randomly-assigned or self-selected groups;
- diversity in the personality attributes further enhances the performance of the group;
- the prediction model can be used to determine the level of performance of a student before actually forming groups;
- the software tool developed can be a viable grouping technique to create effective groups.

In summary, while there are many ways to arrange students to work in cooperative groups, automatic grouping that considers personality attributes and performance level, may be an option. The findings of this research has also provided some useful direction for conducting further research in the areas of education in general and prediction of performance and group composition in particular.

ZUSAMMENFASSUNG

Die vorliegende Studie untersucht Möglichkeiten für die Zusammenstellung heterogener Lerngruppen im Hinblick auf eine Verbesserung des Studienerfolgs und verfolgt dabei im Wesentlichen zwei Ziele: Zum einen geht es um die Entwicklung einer Leistungsvorhersage ohne auf aufwändige Testverfahren zurückgreifen zu müssen und zum anderen um die Entwicklung eines Softwaretools für die Bildung heterogener Lerngruppen.

Die relevanten Variablen zur Erstellung wirkungsvoller Lerngruppen im Anwendungsbereich Mathematik wurden durch Literaturauswertung, Datenerhebung, sowie Diskussionen mit Experten dieses Fachgebietes ermittelt. Die dabei als relevant identifizierten Attribute sind Geschlecht, Gruppenarbeitsverhalten, Interesse für Mathematik, Erfolgsmotivation, Selbstbewusstsein, Schüchternheit, Englischkenntnisse und Mathematikleistung. Es stellte sich außerdem heraus, dass die ersten sieben Attribute auch als Eingabevariable in einem Vorhersage-Modell für das zu erwartende Leistungsvermögen Verwendung finden können. Auf der Grundlage der so identifizierten Attribute wurde ein Fragebogen zur Sammlung experimenteller Daten entwickelt. Die Daten selbst stammen von Elft- und Zwölftklässlern der Senior Secondary School in Addis Abeba, Äthiopien. Basierend auf den experimentellen Daten, wurde ein Bayes'sches Leistungsmodell entwickelt, das eine 70,9-prozentige Vorhersagegenauigkeit besitzt. Durch weitere Experimente sowie eine Modifizierung des Vorhersagemodells konnte die Vorhersagegenauigkeit auf 78,4-prozent gesteigert werden.

Im Hinblick auf die Bildung heterogener Gruppen wurden die Teinehmer der Studie in ein mathematisches Modell abgebildet. Hierfür wurde ein Vektorraummodell verwendet, welches den einzelnen Studenten durch einen Vektor in einem hochdimensionalen Merkmalsraum beschreibt und dadurch die mit dem Studenten verbundenen Persönlichkeitswerte und Leistungsmerkmale widerspiegelt. Auf der Basis des mathematischen Modells wurden drei Algorithmen zur Bildung von Lerngruppen untersucht. Derjenige, der für alle Gruppen eine gleichmäßige Verteilung für die Zugehörigkeit zu den verschiedenen Leistungsniveaus anstrebt, wurde dann für weitere Experimente benutzt. Darüberhinaus befasst sich die Arbeit mit Ansätzen zur Gruppenzusammenstellung aufgrund der Verfügbarkeit studentischer Leistungskennziffern. Ein Ansatz (basierend auf einem off-line-Algorithmus) wird benutzt, wenn ein Bedarf zur Gruppenbildung für eine Menge von Schülern (z.B. Schüler eines Jahrgangs) besteht. Der andere Ansatz (ein inkrementeller Gruppenbildungsalgorithmus) benötigt nicht den kompletten Datensatz im Voraus.

Um das entwickelte Software-Werkzeug zu evaluieren, wurde ein weiteres Experiment mit den gleichen Studierenden der gleichen Schule durchgeführt. Diese Studierenden nahmen zuerst an einem Eingangstest teil. Anschließend wurden sie in drei Sektionen aufgeteilt: Für die erste Sektion erfolgte die Zuordnung zu Gruppen zufällig, Studierende der zweiten Sektion konnten sich aufgrund persönlicher Präferenzen zu Gruppen zusammenschließen und Studierende der dritten Sektion wurden ihren Gruppen durch die entwickelte Software zugeteilt.

Diese drei Sektionen von Studierenden wurden aufgefordert für einen Zeitraum von acht Wochen zusammenzuarbeiten und mussten anschließend an einem Abschlusstest teilnehmen.

Die dabei erhobenen Daten wurden verschiedenen statistischen Tests unterworfen:

- der T-Test für Paardaten, um festzustellen, ob es einen signifikanten Unterschied zwischen Eingangs- und Abschlusstest gibt,
- eine Regressionsanalyse, um die Beziehung zwischen der Gesamtzeit für die Gruppenarbeit und dem Leistungszuwachs zu beschreiben,
- ein Verhältnistest, um die verschiedenen Gruppierungsmethoden zu vergleichen, und
- ein Test für die Differenz der Mittelwerte, um die Stabilität der inkrementellen Version im Vergleich zu dem off-line-Algorithmus zu überprüfen.

Die experimentellen Ergebnisse haben bestätigt, dass

- Studierende, die auf der Grundlage von Leistungsdaten und Persönlichkeitsmerkmalen gruppiert wurden, besser abschnitten, als diejenigen, die zufällig bzw. aufgrund von persönlichen Präferenzen zusammengefasst wurden,
- Unterschiede in den Persönlichkeitsmerkmalen förderlich für die Leistungsfähigkeit einer Gruppe sind,
- das Vorhersagemodell zur Abschätzung der Leistungsfähigkeit von Studierenden im Vorfeld der Gruppenbildung verwendet werden kann und
- das entwickelte Software-Werkzeug eine geeignete Gruppierungstechnik zur Zusammenstellung wirkungsvoller Lerngruppen bereitstellt.

Zusammenfassend kann festgestellt werden, dass trotz der vielfältigen Möglichkeiten zur Zusammenstellung kooperativer Lerngruppen ein automatisches Verfahren auf der Grundlage von Persönlichkeitsmerkmalen und Leistungsindikatoren eine sinnvolle Option darstellt. Die Resultate dieser Arbeit haben auch einige nützliche Hinweise auf mögliche künftige Forschungen im Bereich der Bildung generell, sowie speziell zur Leistungsvorhersage und Gruppenbildung ergeben.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

ACRONYMS

BN	Bayesian Network
CPT	Conditional Probability Table
GAWK	Genetic Algorithm Wrapper for K2
GH	Goodness of Heterogeneity
IT	Information Technology
ITS	Intelligent Tutoring Systems
STAD	Student Teams Achievement Divisions
TPDA	Three Phase Dependency Analysis

CHAPTER ONE
1. INTRODUCTION

1.1 A Brief Overview of Cooperative Learning

Cooperative learning is defined as the instructional use of small groups to help students work together and maximize their own and one another's learning, develop a positive image of self and others, provide vehicle for critical thinking and problem solving and to encourage collaborative social skills (Calderon, 1987 as quoted in Christison, 1994; Johnson and Johnson, 1990a).

Groups may be formed with various learning objectives. They include **skill exercises**, where students demonstrate their understanding of a topic; **guided discovery learning** where students learn through discovery rather than being told the information directly; **in-class problem solving** where instructors allow students to practice problem solving; and **long-term problem solving project** where students are required to carry out a term- or semester- project with careful planning (Apple, 2001). The focus of this research work is the case of in-class problem solving.

In the literature, various names have been used to refer to cooperative learning: collaborative learning, collective learning, learning communities, group learning, study groups, reciprocal learning, team learning, etc. (Davis, 1993). The terms "cooperative learning" and "group learning" are used interchangeably in this research work.

Cooperative learning has been one of the many alternative instructional techniques described in the academic literature to enhance student performance (Dansereau and Johnson, 1994; O'Donnell and Dansereau, 1992; Webb, 1992; Johnson and Johnson, 1989; Slavin, 1983a, 1996). The consensus among workers is that performance on a subject is enhanced when an individual learns information with others as opposed to when she or he studies alone. Students who work in groups are observed to develop an increased ability to solve problems, show greater understanding of the subject being taught, and retain it longer than when the same content is presented in other instructional formats (such as individualized instruction). Further, several works reported that students were more likely to acquire critical thinking skills and meta-cognitive learning strategies, such as learning how to learn, in small group cooperative settings as opposed to listening to lectures (Slavin, 1991; Johnson and Johnson, 1990b; McKeachie,1986; Dishon and O'Leary, 1984).

Cooperative learning has also proved useful in large class size environments. For instance, Hake (1998) stated that students in large class size, who attended lecture based courses, had consistently lower performance than those attending classes characterized by some form of active learning methods, such as group learning.

Apart from academic achievements, the social benefits that accrue to students from cooperative learning experiences have also been well documented by Jordan and Le Metais (1997), Slavin(1995), and Kamps et al(1994). As such, cooperative learning is widely recognized as a teaching strategy that promotes learning and socialization (Cohen, 1994).

According to Johnson and Johnson (1990a) and Cowie et al (1994), cooperative learning is more than:

- having students sit side by side at the same table and talk with each other as they do their individual assignments, or
- having students do a task individually with instructions that the ones who finish first are to help the slower students, or
- assigning a report to a group where one student does all the work and others put their name on it.

Rather, it is about encouraging or requiring students to take turns in helping one another learn through discussion on subject matters, reading and reviewing course materials, completing course assignments, commenting on each other's written work, preparing for tests and exams, helping each other with difficulties that are encountered in class, etc. Accordingly, the essence of cooperative learning is more in the creation of a learning environment that:

- brings positive interdependence among team members;
- helps students develop a sense of shared community where group members transcend the gender, racial, cultural and other differences they may sense among themselves through the introduction of high-level group communication and face-to-face interaction;
- helps students get emotional and academic support and persevere against the many obstacles they face in school.

Cooperative learning can be used in all subject areas. Particular to Mathematics, studies in the area consistently prove that there are many benefits from using a cooperative learning approach in the Mathematics classroom. More is said on this in Section 2.4 of Chapter 2.

In the case of Ethiopian educational environment, where student enrolment at various levels is increasing, it is observed time and again that students appear disengaged during lectures, demonstrate low levels of understanding and get low grades. As such, it is envisaged that learning and discussing in small groups might help students to retain more than the conventional classroom lecture.

On the whole, in cooperatively structured learning of any subject area, students work together to attain group goals that cannot be obtained by working alone. This also results in more interpersonal relations and greater support regardless of the differences in ability.

1.2 The Problem

Although the advantages of cooperative learning are well documented, group productivity or improvement in individual performance is very much determined by how well members work together. The educational benefit that a learner gets through group learning depends mainly on interaction among the learners (Cowie et al, 1994; Johnson and Johnson, 1990b; Jacobs, 1988; Ames and Ames, 1985). Further more, Slavin (1987) and Johnson and Johnson (1985) claimed that many of the unsuccessful outcomes from group work stem from the composition process. Such factors as determining the size of a group based on the learning objectives and structuring of lessons cooperatively are important in the formation process. However, it is the composition of group members (the allocation of students into groups) that takes into account inter-working ability among members, which seems to be important in forming effective groups.

Although, there is no "one right way" to allocate students into groups, there exist a number of practices in use. Most of the practices, however, depend heavily on forming the groups based on ability or performance level of each student in the class. For instance, professor-formed groups based on pre-test scores are common.

In an attempt to address the issue of allocating members into groups, Slavin (1987) proposed that in addition to assigning students into groups randomly, or allowing them to create their own groups, they should work in small, mixed-ability groups of four members: one high achiever, two average achievers, and one low achiever. Even further, it is argued to consider, in addition to ability, personality and non-academic attributes[1], that help members interact better. Studies conducted by Bradley and Herbert (1997) and O'Donnell and Dansereau (1992) also emphasized the importance of personality attributes (personal and social characteristics) in group composition. According to these authors, personality attributes determine whether the groups perform according to the desired goals or attainment. Other researchers in the area (Martin and Paredes, 2004; and Romney, 1996) suggested that in addition to performance levels, attributes such as gender, family and school background of the student, instructional language proficiency, ethnic background, motivations, attitudes, interests, and personality (argumentative, extrovert, introvert, etc.), should be given due attention in the process of forming groups.

It is also observed that although homogeneous groups are better at achieving specific aims, when students with different abilities, experience, interests and personalities were combined (heterogeneous groups), they out-performed homogeneous groups in a broader range of tasks (Martin and Paredes, 2004; Nijstad and Carsten, 2002). Heterogeneous grouping works with the assumption that groups work better when the members are balanced in terms of diversity based on functional roles or personality differences. In other words, students in effective groups should be diverse in backgrounds, ideas, personalities, ethnicity, and gender (Slavin, 1995; Romney, 1996).

In view of the foregoing, an issue that is gaining more and more popularity among workers in the field of cooperative learning is the formation of heterogeneous groups (whenever possible). This is done based on a set of specific criteria[2] applicable for the learning objective under consideration. The implementation of the task usually involves students completing a questionnaire which is scored to determine a student's personality characteristics. Students with different performance levels and characteristics are then appointed to each group so as to achieve the desired balance in terms of diversity (Interactive Media and Learning, 2005)[3].

While generally considered very effective, such a task is not without its drawbacks especially in large size classes. Since it requires questionnaires to be developed, administered and scored (all prior to the group formation), it can be expensive and time consuming. Moreover, in a manual environment, a great deal of time and effort may be needed in the

---

[1] The terms Traits and attributes are interchangeably used in this thesis.
[2] See Chapter 2 for details of the specific criteria method in group formation.
[3] **Institute for Interactive media and learning**
http://www.iml.uts.edu.au/learnteach/enhance/groupwork/unit3.html
(last visited August 20, 2005)

creation of heterogeneous groups with the required features. This is because, the numbers and combinations of performance level and values of personality attributes to be considered may be too many to handle and manage. In addition, where most students have different abilities and styles, the application of this grouping method using the simple models developed by workers in the area so far may lead to an over-representation of some styles and under-representation of others. In the absence of effective practical models that fully consider the group formation with features incorporating performance and personality attributes, difficulties may be experienced in the realization of the benefits from heterogeneity. Hence, the creation of effective heterogeneous groups may be difficult using the existing (those that are being explored so far) simple and manual methods.

Nowadays, computer-based tools in the areas of education, to support both individualized and collaborative learning, are very popular (Krejins, 2002; Florea,1999; Bull, 1998; Chan et al, 1995; McConnell, 1994; Collins and Brown, 1988). However, no serious attempts are being made to extend the applications of computers to maximize benefits from collaborative learning by automating the group formation. Such attempts were observed to be lacking from workers in both the educational and social psychology and computer science fields. One of several reasons for this could be the lack of exposure (on the part of educational psychologist) to the potentials of computers to capture knowledge and reason in such problem domains that involve the study of socio-cultural attributes. While there is considerable effort to employ computer-based tools for educational applications by educationists, most of the efforts seem to be in the use of conventional methods (those designed for use in solving structured problems) than the emerging innovative methods applicable in the unstructured domains. On the part of computer scientists as well, it is not difficult to observe the limited exposure and awareness of interdisciplinary research opportunities in the challenging fields of educational and social psychology and computer science. The above notwithstanding, more recently, encouraging joint efforts are being made elsewhere, to investigate the applicability of such emerging computer methods in the areas of adaptive issues in e-learning systems, intelligent tutoring systems, student modelling with dynamic information updates, etc.

Taken together, despite the critical role that group formation is said to have in terms of enhancing the effectiveness of cooperative learning on the one hand and the potentials of computer-based methods to facilitate (assist in) the group formation process on the other, no serious research work is reported that addresses this area on its own or even as part of the work in the area of student modelling (a related area that is being widely explored).

Although the consideration of both academic performance and personality traits for group formation has been widely advocated as ideal and beneficial in terms of enhancing the effectiveness of collaborative learning, it has not been explored/developed fully and thoroughly in practical terms. Most of the attempts so far reported were either anecdotal/sketchy or limited in scope and purpose. What is more, there are generally lack of publicly available software tools (whose development is based on proven models and methods) for use in the formation of groups that enhance cooperative learning. Hence, the motivation for this research.

As such, this research work was an attempt to explore possibilities of developing a mathematical model good enough to consider both performance and personality traits in the formation of effective heterogeneous groups; and to develop an appropriate software tool to implement the model (i.e., automatically create groups with the help of computers).

As a result of preliminary investigations for the purpose of identifying the most critical aspects of the study, the following specific areas of work have been identified:

- Identification of personality attributes to be considered in the composition of heterogeneous groups;
- Building and then programming a mathematical model (based on the attributes identified) that will assist in the creation of heterogeneous groups automatically;
- A machine learning (Bayesian network) based prediction model to determine the performance level of students to be assigned to the groups.

While the identification of the first two areas of the study may be obvious from the preceding discussions, the reasons behind the identification of the Bayesian network based prediction model as one of the specific areas of study may deserve the following explanations.

With regard to the personality attributes to be considered in group composition, once they are identified, their corresponding values may be obtained directly from the student in the form of responses to questions, from student fact sheets or through surveys conducted for this purpose. With regard to determining academic performance level of students, however, the most commonly used technique has been the administration of exams which may be expensive and an extra work for the instructor, particularly with large class sizes and when done for the sole purpose of group formation. What is more, the appropriateness of obtaining performance level using exams only is being criticized and challenged by workers in the area (Humphreys et al, 1982).

Under the circumstances, therefore, it was considered helpful to devise an automatic model that predicts performance based on known information and without necessarily requiring students to write exams. In relation to the information required to predict performance, a number of research studies have identified factors that intervene with performance level of students (Diaz, 2003; Caplan, 2002; Atkinson, 2000; Sewnet, 1995; Daniel, 1992; Mekonnen and Darge, 1991). Among the elements identified to intervene with student performance are individual differences (personal causal factor); parents (family causal factors); teaching/learning strategy (system causal factor); and teachers (academic causal factor). Some even conducted statistical analysis to explain the relation between the various factors and performance level (Daniel, 1992; Demis, 1991; Taddese, 1990).

As mentioned earlier, it may be possible to find information on most of these factors directly from students (at the time of forming groups) in the form of responses to questions, answers to problems, or from student fact sheets, and apply some sort of association rules to predict performance manually. In reality, however, this task is far more complex than what it does seem to be. For instance, there are various potential sources of errors in this mode of information gathering that may affect the resulting performance prediction. First, the student's responses may be affected by lack of seriousness and honesty, arising from inter- and intra-individual variability to mistaken self-concept. One can never be certain that the data is truly representative of the current state of the student. Second, since it is obviously impossible to know everything about a student, manually establishing appropriate association rules to efficiently determine the likely performance is not a simple task. If the rules are inconsistent, incomplete or semantically inexplicable, the quality of the data will have a reduced bearing on the quality of the resulting prediction.

Even the relationships and dependencies between personality attributes themselves and to performance are very complex and involve unstructured and uncertain-reasoning dimensions. Often, complex relationships arise in the process of prediction with increased number of attributes, and such complexities lead to uncertainty about the basis for observed relationships and dependencies. Moreover, as presented in Chapter 2 of this thesis, what brings further complication to the issue is the fact that even published literature revealed conflicting results on the significance of some factors affecting performance. For instance, while some claim that gender and educational background of families significantly affect performance, others do not agree (Comelius and Cockburn, 1978; Sewnet, 1995; Khandker, 1996; Diaz, 2003; Bedilu, 1996; Entwistle, 1972).

All these make the modelling of the prediction very challenging and even difficult to model using conventional methods based on exact inference and reasoning. It is believed that recent advances in knowledge representation and reasoning (learning and inference methods) offer valuable tools for dealing with such uncertain problem domains. And hence it is important to explore the application of machine learning methods such as Bayesian Networks (BN) for the development of a model to predict the performance level of students to be assigned to groups. A Bayesian Network, as detailed in Chapter 3, is a powerful knowledge representation and reasoning tool for expressing what one is certain and uncertain about (Jensen, 1996).

In what follows, the research questions that arose from the discussion of the problem are presented.

1.3 Research Questions

Based on the foregoing discussions, the following major questions guided the research work:

Identification of attributes:

(i) What are the major attributes/characteristics of students that may be considered in group composition?

(ii) What are the major factors that affect the level of performance?

(iii) Which attributes are common both in group composition and performance prediction?

(iv) What is the technique for acquiring values for those identified attributes?

Performance prediction model:

(v) How can one build the Bayesian network based on these attributes?

(vi) Would performance predicted by the software tool be as good as performance based on actual exam results?

Heterogeneous group composition:

(vii) How can one build and program a mathematical model to automate the formation of heterogeneous groups?

(viii) Would there be an improvement in performance of students as a result of group work?

(ix) Would the automated group composition increase performance of students as compared to the other grouping methods?

1.4 Methodology/Approaches

This section briefly introduces the experiments carried out in order to address the above research questions. Detailed accounts of the experiments conducted are presented in Chapters 4, 5 and 6 of this thesis.

(i) Exploration of the problem

- In order to define the research problem properly, several preliminary interviews were conducted with individuals in the field of Educational and Social Psychology;
- Relevant literature on group composition, performance factors and Bayesian networks were reviewed.

(ii) Selection of Subject Area

- To contextualize the experiment to be undertaken, mathematics was taken as the subject area. This was mainly because of its vital importance in the school curriculum. Education systems throughout the world place high importance on the teaching and learning of mathematics and a lot of effort is being made to improve efficiency and effectiveness in these activities (Garden, 1987). See Section 4.1.3 of Chapter 4 for further details.

(iii) Selection of Attributes

- In order to reduce the complexity of the work being done in terms of identifying the most relevant performance factors as well as the ones to be considered in group composition, it was important to identify the common attributes to address both issues;
- Review of literature, formal and informal interviews were conducted with mathematics teachers, educationists, and students at Addis Ababa University in order to understand the current situation of group work and performance in mathematics tests. This phase helped for initial identification of attributes for consideration in the process of group formation;
- In order to make final selection of attributes, a survey was conducted. This survey helped to identify the attributes that may be considered in group composition. In this process, attributes that intervene with performance were also identified.

(iv) Identification of Target Data Sets

- The test participants were students of one senior high school in Addis Ababa who were in their final year of preparatory[4] program to join higher learning institutions.
- These students were asked for their consent to participate in the experiment with persuasion from the mathematics instructors of the school.
- Issues of data protection and privacy were also discussed with students before the actual data collection took place. More is said on the above two issues in Section 4.1.2 of Chapter 4.

(v) Collection and Preparation of Data

- Once the attributes were identified, instruments for the purpose of data collection were developed and tested in consultation with educational psychologists. Details are given in Section 4.3 of Chapter 4;
- Data have then been collected and processed for use both by a Bayesian network software and in the group composition experiment.

(vi) Building the Performance Prediction Models

- As detailed in Chapter 5, in the experiments conducted to build the prediction model, a Bayesian Network Tool in Java(BNJ)[5] and the Bayesian Network PowerConstructor[6] software tools were used for reasons of accessibility;
- The Three Phase Dependency Analysis (TPDA) algorithm was employed to learn dependencies and relationships between the attributes considered (i.e., the structure of the network);
- The Lauritzen-Spiegelhalter exact inference algorithm, available in BNJ, was employed to predict performance. Further explanations on the algorithms are given in Chapter 3.

(vii) Testing the Prediction Model

- The performance of the Bayesian network model was tested as follows:
    - o Mathematics exam was administered to selected students from the same high school and the performance of each student was recorded;
    - o The same students were made to fill out the attribute measuring instrument and the performance of each student was predicted;

---

[4] As of 2003/2004, preparatory program in high schools is equivalent to Freshman program in the University.
[5] http://sourceforge.net/projects/bnj/
[6] http://www.cs.ualberta.ca/~jcheng/bnpc.htm

- The predicted performance of each student was compared with the corresponding actual performance. (Details are given in Section 5.2 of Chapter 5).

(viii) Forming Heterogeneous Groups

- A mathematical model, that addressed the group formation problem, was developed.
- Three algorithms were developed, based on the mathematical model, to form learning groups and the corresponding Java programs were written;
- The algorithm which generated the best group composition in terms of heterogeneity was selected for further grouping of students.
- An attempt was also made to compose groups on an incremental basis, i.e., an incremental version of the selected algorithm was developed for this purpose. Details of all the experiments on forming groups are provided in Chapter 6.

(ix) Testing the Automatic Group Composition

- Students from the three sections, who were participants in testing the accuracy of the prediction model, were made to study in groups. While students in one section were grouped randomly, students in the second section were allowed to form their own groups. Students in the third section were grouped automatically by the program developed;
- Post-group work exam was administered;
- Statistical tests were applied to check whether there was a difference in extent of heterogeneity within groups when using the batch processing algorithm or the incremental version (See Section 6.5 of Chapter 6).

1.5 Contributions

In view of addressing the research questions listed above and with the general objective of forming heterogeneous groups, the following may be considered as major contributions of the research work.

- A mathematical model that addressed the group formation problem in cooperative learning, through the mapping of both performance and personality attributes into a student vector space, was developed. This served as a foundation for the application of formal methods in determination of heterogeneous groups based on both performance and personality attributes. The mathematical modelling approach introduced as part of this work, which is original in this area, may be explored further in such related research undertakings as collaborative learning and student modelling.

- A portable software tool (java programs) to implement the grouping model was developed.
- A machine learning (Bayesian Network) based performance prediction model was developed. This model helps in predicting the individual performance of a student based on known information, without necessarily requiring the student to write exams. In this study, the predicted performance value is used as one of the factors in determining the group to which a student is to be assigned to benefit from cooperative learning. It is, however, relevant to note the fact that practitioners and workers may also use the model to predict performance of a student, particularly where exam administrations become difficult or impossible.
- The claim that grouping based on profiles of students has a potential of improving performance, is confirmed. Researchers in the area of cooperative learning can make use of the results of this study in order to make optimal and effective group composition.
- The result of this research work may be an insight into extending the student model component of an intelligent tutoring system to include personality attributes in addition to subject matter knowledge. To this end, a software tool might be developed which interacts with the student model to suggest grouping possibilities.

In general, this work may be considered a special contribution in terms of providing some useful direction for conducting further research in the area of predicting student performance and group composition to enhance the effectiveness of cooperative learning, particularly, in the context of the Ethiopian educational environment. The work is not only the first of its kind in Ethiopia in terms of applying cooperative learning techniques to improve the performance and relationships of students in the mathematics subject, but also in the creation of opportunities (through the formulation and setting of the research) for computer science, educational psychology and social psychology researchers to jointly contribute to (or participate in) such multidisciplinary area of work.

1.6 Limitations

The following are the major limitations of the study:

- In spite of the persuasion and encouragement both from the teachers in the school and the researcher, some students were reluctant to provide information about their personal behaviours. Some did not even complete the survey instrument. Survey instruments on which almost all answers to the questions were incomplete or answers exaggerated were discarded.
- Because of administrative and policy related issues, it was a challenge to collect the required size of data, administer exams and arrange the time for group works. The alternative approach was to spend more days than actually necessary in collection of data, exam administration and duration of group work. These have made the time needed for the experiment longer than expected.
- Considering the various constraints in this research, only eight personality attributes were considered in the experimental work of predicting performance. Inclusion of more performance factors might lead to better prediction accuracy.

1.7 Organization of the Thesis

This thesis is organized into 8 chapters. Chapter 2 provides a review of literature related to performance factors and cooperative learning. Chapter 3 introduces basic probability concepts followed by a detailed description of learning and inference in Bayesian networks. This chapter also briefly presents application of Bayesian networks in the field of education. Chapter 4 presents the survey works related to identification of attributes, development of instruments and measurement of the attributes identified. Chapter 5 deals with experimental works related to development of the performance prediction model. The prediction accuracy of the developed model is also presented in this Chapter. Chapter 6 deals with the experiments on forming heterogeneous groups. The design of the algorithms and their core features are described. Experiments carried out on three grouping methods including the automatic group formation are also presented in this chapter. Experimental results and discussions, in relation to the research questions posed at the beginning, are presented in Chapter 7. Chapter 8 ends in giving concluding remarks as well as directions for future work.

CHAPTER TWO

2. PERFORMANCE FACTORS AND COOPERATIVE LEARNING

As a background to the survey conducted in selecting relevant attributes for the study, this chapter attempts to highlight issues related to performance factors and group composition. Factors intervening with performance are presented first, followed by further background (to those given in the previous chapter) on cooperative learning and group formation. After a brief remark on the benefits of cooperative learning in mathematics teaching, the chapter concludes with a summary. It is to be noted that mathematics is the subject matter selected for experimentation in this study.

2.1 Factors Intervening with Performance

Many Educators and Psychologists have long been concerned with understanding the factors that contribute to the differences among students in relation to academic performance. Some researchers have also come up with explanations on some of the factors. Based on the available literature, this section presents some of the academic and non academic factors frequently cited in the literature to explain the performance level of students.

2.1.1 Academic / School Related Factors

Lack of School Materials and Facilities: According to Stromquist (1997), Schiefelbein et al (1994) and Fuller (1987), availability of educational materials (text books, school library, laboratory, etc.) significantly affects academic performance of students. It is also a widely accepted fact that schools with better facilities and materials to facilitate the instructional process are possibly more efficient than others without such facilities. Thus, together with other factors, the scarcity of school materials (textbooks, reference books, etc.) that are particularly related to instructional activities affect the educational performance of students (Adane, 1993).

Difficulty of Instructional Language: Lack of familiarity with the language of instruction is an obstacle for students coming from a deprived background and this difficulty is felt with particular acuteness with regard to academic failure (Gall et al, 1973).

In the case of Ethiopia, although English is taught as a subject beginning in grade one and used as a medium of instruction starting from the seventh grade, it is not adequate enough to enable students to easily understand spoken as well as a written text in the English language.

Researchers in this area indicated that students had difficulty in understanding and using English. For instance, Darge (1989) said that many students have problems with the English language, as this is not their native language. As a result, they might find it difficult to follow lectures, to understand their study material and to take notes in English. In addition, it is not easy for such students to ask or answer questions. A study by Zaudneh, Darge and Nardos (1989) also revealed that 58% of a sample of first year students of Addis Ababa University had difficulty in understanding lectures in English. According to those researchers, the language of text books may be too difficult for students to understand.

Motivation to Learn: Research findings indicate that achievement motivation and academic performance are correlated. According to Zsolnai (2002), students with high achievement motivation perform faster and more competently than students with low achievement motivation. As such, motivation maintains a relationship with the level of hard work and this, in turn, with performance (Nunez et al, 1998 cited in Diaz, 2003).

Similarly, Corno (1986) and Zimmerman and Martinez-Pons (1986, 1988) proposed students' motivation on classroom academic tasks as one component (factor) for academic performance. For instance, capable students who persist at a difficult task or block out destructors, such as noisy classmates, maintain their cognitive engagement in the task enabling them to perform better.

Interest in the Subject: It is observed in various research studies that low achievers are the ones who hold negative views towards school-oriented activities. According to Raph et al (1966), low achievers professed disliking for their course and professors. They reported being more easily discouraged when confronted with long or difficult assignments, being accustomed to exerting only minimum effort in courses which they did not like.

Peer relations: Interactions with peers promote acquisition of social competencies (Marchesi and Martin, 2002 cited in Diaz, 2003). Studies by Montero (1990) cited in Diaz (2003) also proved that positive correlations existed between performance and peer relationships. It was demonstrated that students failing in school were those most rejected by their groups or classmates.

2.1.2 Non-Academic Factors

Gender Differences: Good and Brophy (1990) indicated significant gender differences in patterns of motivation and achievement in various subject-matter areas. They stated that males tended to score higher on tests of visual spatial ability and mathematical ability, and females to score higher on tests of verbal ability. Wilkons and Marrette (1985) also indicated that mathematics test-score patterns favoured girls in the early grades and tests of number computations and mathematical reasoning favoured boys in the latter grades.

Other research outcomes on gender differences in academic abilities concluded that boys and girls showed equal aptitude and achievement in arithmetic until they were well into the elementary school period (Maccoby and Jacklin, 1981; Fenneman, 1974). It has further been noted that after the fourth grade,achievements tend to be in favour of boys.

Although girls and boys were not different in mathematics background or performance at primary levels, significant differences favouring males were found at secondary levels (Sherman, 1980). However, a study by Diaz (2003) explained gender differences as part of the variation in academic performance $F(1,122) = 14.89$, $p<0.001$, where girls showed better performance.

In the Ethiopian context, boys seem to perform better than girls at all grade levels (Tsige, 1991; Ademe and Gebre, 1990). This may be due to the fact that boys were more favoured than girls in terms of access to education. Pervasive gender ideologies at the household and community levels often favour males over females and thus promote differential education opportunities and outcomes. Girls' work requirements for the family is heavier than boys. They would often be required to work from an early age in the household occupying most of their time and, in due course, they do not have time to study at home resulting in their low academic performance.

Age: Based on his research findings, High (1996) claimed that age is a good predictor of performance of a student. Mathewos (1995) also reported that age significantly affects the achievements of students. Although maturity may be positively correlated with performance, this may not be the case in older ages. For instance, in a study by Diaz (2003), age was proved to be important in explaining academic performance, $F(1,122) = 263.05$, $P < 0.0001$, showing that among older students there are more repeaters. Age appeared as an explaining variable in the affective-motivational aspects: as a student gets older, the scores observed for academic performance decreases.

Economic Status of Parents: Many studies have found that the economic status of parents is a significant factor for poor academic performance of students. According to Comelius and Cockburn (1978), low performing students are mostly from poor families. It appears true that economically better off parents would be able to provide their children with the necessary learning facilities than poor parents. Although this may be mostly the case, there are also instances that indicate otherwise. For instance, the research results by Sewnet (1995) revealed that students from poor families showed better academic performance than those from well off families. In general, one cannot dispute the fact that well off families are better positioned to positively contribute to the academic performance of their children by way of providing required facilities and support. On the other hand, the positive contribution to performance that indirectly results from the determination and commitment of the students from poor families (to improve the economic situation of their families as a result of their education) cannot be underestimated. Hence, economic situation of parents may be considered as a factor either way, although the scale and degree of influence might be contextual.

Educational Background of Parents: Educational background of parents is an important determinant factor for academic performance of students, i.e., students from illiterate parents perform lower than those from literate ones (Carron and Chau, 1996; Akinkugbe, 1994, Magland, 1994).

Even though many scholars agree on the opinion that educational background of parents affects academic performance, there are differences in the view that the literacy status of fathers and mothers equally affects academic achievement. For instance, from research findings of Khandker (1996) and Gill (1991), one observes that education of mothers is more closely associated with the academic achievement of daughters than is the education of fathers. The study of Diaz (2003), however, revealed that the only factor with explicative ability of academic failure is education of fathers, $F(1,109) = 3.454$, $P<0.05$. Students whose fathers have higher-level studies were those who least failed.

Researchers in Ethiopia further claimed that education of mothers increased academic performance of girls (Rose et al, 1997 and Yelfign et al, 1995). Although it is observed in various research studies that the educational level of parents could influence the scholastic achievement of their children, there are also instances in which family education level has nothing to do with achievement. According to a study made by Sewnet (1995), high school students from illiterate families scored higher on mathematics test than those from families with elementary or high school background. Taking into account the sample size taken, it may be difficult to generalize for the whole population of Ethiopian students. Most rural parents with little educational background may hardly provide with the necessary encouragement and incentives for schooling, resulting in low academic performance of their children.

Parenting Styles: According to Baumrind (1991), parents high in both love and control are considered authoritative. Parents low in both are considered to be neglectful. Those high in love and low in control are considered to be indulgent and those low in love and high in control are said to be authoritarian.

There are a few general preliminary studies which provide some insights about the issue of parenting practices in Ethiopia and impacts on academic performance of children. Markos (1996), in his study on the relationship between parenting style and school performance among high school students in Mekelle, found that parenting style had a significant contribution to the school performance of the students. According to his report, high school students who characterized their parents as authoritative achieved higher in school than their counterparts who described their parents as authoritarian, indulgent or neglectful.

The study conducted by Birhanu (1996) on the relationship of parenting styles with academic achievement among senior secondary school students in one ethnic group (Keffecho zone) also found out that authoritative parenting style was positively related to academic performance. While authoritarian parenting style was negatively so, significant relations were not observed between either of indulgent or neglectful parenting style and academic achievement.

Cox (1967) with 137 university students indicated that 65% of the students believed that their parents had been too strict and had not allowed them enough freedom. He reported that they have to accept rules that were established by parents with out argument and that they experienced excessive control in the home, which resulted in low academic performance in school. Similarly, Haile (1970) stated that children in Ethiopia were culturally restricted not to exercise self assertion, self-esteem, and self- reliance. He indicated that parents defined their position by invoking their religious belief that a child should be obedient and should comply with the wishes of elders. A study by Haile also implied that the parenting style dominantly practiced in Ethiopia was authoritarian.

In general, it should be noted that the type of parenting style dominantly practiced in Ethiopia is authoritarian.

Individual Differences: In describing the role of personality attributes Davis (1971) discussed that student's perception of who s/he is, what s/he enjoys and what s/he can do or can not do in the learning situation affects her/his responses to the ongoing learning. This in turn affects her/his performance on academic activities.

Research findings by Entwistle (1972) and Elliot (1972) as cited by Anthony (1973) indicated that students with extroversion characteristics were found better in their performance than introverts. Bedilu (1996) compared the average scores of introvert and extrovert students in three Ethiopian colleges in reading and writing tests in Amharic. He found no significant difference. However, introverts scored better than extroverts on the writing test.

Whilhite (1990) examined the relationship of study behaviours and academic achievement of students. He found that scores on self assessment measure of memory ability and scores on a locus of control measure were the best predictors of final course grades. Diaz (2003) also found out that 34% of the variation in academic performance, in a sample of 1178 secondary school students in Spain, is due to personal attributes such as age, academic self-concept and locus of control (self-confidence).

This completes the brief review of literature relating to factors intervening with performance. As mentioned earlier, some of the performance factors presented in this section formed the basis for the survey works reported in Chapter 4.

2.2 Cooperative Learning

2.2.1 Nature and Psychological Basis of Cooperative Learning

Slavin (1995) defined cooperative learning as a set of instructional methods in which students are encouraged or required to work together on academic tasks. He noted that such methods might include having students sit together for discussion, or help each other with assignments and more complex requirements.

According to Kagan (1994), there are four basic principles which define cooperative learning: Positive Interdependence, Individual Accountability, Equal Participation, and Simultaneous Interaction. Kagan uses the acronym PIES to represent these principles, and asserts that unless all of these principles are implemented, it is difficult to say cooperative learning is taking place. A brief overview of these principles is felt in order.

Positive Interdependence means that a gain for one student is associated with gains for the other students. As such, for effective cooperative learning, it is important that students are guided to understand that, "the success of every team member depends upon the success of each of the other members", and "if one fails, they all do." (Kagan, 1994). In fact, this "sink-or-swim-together" mentality is the central theme of cooperative learning. One way to foster positive interdependence, for instance, is to not give each group member all of the necessary materials to complete the assigned task, thereby forcing the students to share and work together.

Individual Accountability means that each group member is responsible for his or her own learning, and for contributing to the learning of his or her group members. What this means with regard to grading, for instance, is if the teacher is to assign a group grade, it is also important to assign individual grades to each student, based on exams and other work which is done independently.

Equal participation refers to the fact that no student should be allowed to dominate a group, either socially or academically, and that no student should be allowed to be idle, or "hitchhike" on the work of other group members. Kagan cautions that equal participation does not occur automatically, and that steps must be taken to ensure that it occurs. Among the techniques recommended to ensure equal participation are the following. One is turn allocation, which means that students are expected to take turns speaking, and to contribute to the discussion when their turn comes. Another one is division of labour, which means that each group member is assigned a specific role to play in the group.

Simultaneous interaction in cooperative learning results from arranging the students in small groups, seating the students face-to-face, and creating a group task such that all group members need to work together to obtain a solution. This could be contrasted to a traditional classroom setting in which all of the students are facing forward, working independently, and spending the overwhelming majority of the time sitting quietly, listening only to the teacher.

According to Apple (2001), while cooperative learning can take different forms and be implemented in numerous different contexts, it is distinguished by the following essential components:

- Groups are formed based on the learning objectives and a set of pre-determined criteria;

- Mutual (positive) interdependence develops when a student believes that he or she can not survive alone and the entire group is required for success;
- With high-level communication and face-to-face interaction, students help, assist, encourage, and support each other's efforts to learn;
- With Inter- and intra- group teaching, there exists much greater student interaction and the possibility for students to teach each other;
- Each individual is accountable for his or her own learning;
- Students can do ongoing reflection and assessment so as to improve future performance;
- Students develop a sense of shared community among members as a result of camaraderie, respect, social cohesion and bonding.

The many features of cooperative learning are supported by extensive research, and are grounded in the theories of Educational Psychologists. As such, cooperative learning is supported by both cognitive and non-cognitive (Behavioural and Humanistic) theories. In Rothstein (1990) and Ormord (2003), it is stated that cooperative learning provides for the application of such **cognitive learning principles** as getting the students involved and making time for over learning through repeated practice sessions. The **humanistic learning theory** is very much related to cooperative learning, since it allows students to make choices in the learning process and explore their feelings and emotions through small group discussion. It encourages students to empathize with other students by discouraging stiff competition/rivalry. Moreover, the application of such **behavioural learning principles** as making learning tasks fun or pleasant experience by minimizing competition, not forcing all students to progress at the same pace, and encouraging more active participation are characterized by cooperative learning. It is also indicated that cooperative learning promotes the application of **social learning theories** through letting students think aloud and praising desired behaviours during group activity. In fact, cooperative learning seems to be strongly supported by social learning theory. According to Bandura (1977), social learning theory considers that students learn from one another, including such concepts as observational learning, imitation, and modelling. Students can learn by observing the behaviours of others and the outcomes of those behaviours.

2.2.2 Benefits of Cooperative Learning: Theory and Research

Before going to details on empirical research on cooperative learning, it may be important to highlight the distinctions between competitive, individualistic and cooperative learning.

In a **competitively structured** classroom, students engage in a win-lose struggle in an effort to determine who is best (Johnson and Johnson, 1991). Students perceive that they can obtain their goals only if the other students in the class fail to obtain their own goals (Johnson, et al, 1986). They are especially concerned about outperforming their classmates. Students in **individually structured** classrooms work by themselves to accomplish goals unrelated to those of the other students (Johnson and Johnson, 1991). In a **cooperatively structured** classroom, students work together to attain group goals that cannot be obtained by working alone or competitively. In this classroom structure, students discuss subject matter, study together, read and review course material, complete course assignments, comment on each other's written work, prepare for tests and exams and help each other with difficulties that are encountered in class.

Johnson and Johnson (1994) stated that although individualistic and competitive teaching/learning methods certainly have their place in the instructional program, they should be balanced with cooperative learning. Slavin (1995) affirmed that with the appropriate use of cooperative learning, improvements were shown in relation to student achievement, ethnic relations, acceptance of academically handicapped students, and self-esteem. Johnson and Johnson (1990a) further indicated that cooperative learning experiences promoted higher achievement and productivity as compared with competitive or individualized learning. Supporting the above idea, Williams and Burden (1997) claimed that in cooperative learning, students encouraged and facilitated each other's efforts to achieve. Christison (1994) also put forward the results of many studies showing that all high, average and low achievers gained from a cooperative experience.

Baris-Sanders (1997) wrote about her experiences in cooperative learning as follows:

> "I was unsure of what would happen when I paired Tatsuo, a very active student, with a shy student, whom I had never heard speak before. After completing an exercise in English conversation in pairs, I was happy to see them volunteer to demonstrate for the class. The students in the class knew that Tatsuo was ready to talk, but they, too had never heard Sohei speak before, to say nothing of seeing him standing in front of them. They watched wide-eyed and open-mouthed until the two had finished their

> conversation. Then they clapped and exclaimed to one another in amazement. Sometimes having a partner can make all the difference".

Similarly, scholars from different backgrounds and disciplines have validated the results of cooperative learning using studies in different contexts and settings with all types of students (varying age, sex, class, nationality, and cultural background). Students who work in collaborative groups also appear more satisfied with their classes (Goodsell et al, 1992; Chickering and Gamson, 1991; Beckman, 1990; Cooper, 1990; Collier, 1980).

Several experiments (Florea, 1999; Humphreys et al, 1982; Allen and Van Sickle, 1984; Johnson and Ahlgren, 1976) also confirmed the effectiveness of learning groups especially in large size classes. Since members in a group may have different ways of explaining the same topic, a student may gain more from a learning group than from other methods of learning. This is also guided by a notion that students can often do as a group, what they can not do by themselves and that students can benefit from peer teaching/explanations. Although it is not possible to cover as much material during the semester as is done in lecture, experiments show that students who work in groups develop an increased ability to solve problems and understand the material.

Bryan (1996), in his article "Cooperative Writing Groups in Community College", stated that

> "In a writing course, cooperative writing groups are very effective because students are more actively engaged in the content of the course; establish a supportive, comfortable learning environment; and experience greater gains in mastering course content."

In the area of cooperative groups in language learning, Holt (1993), McGroarty (1991) and Swain (1985) further confirmed that cooperative groups increased opportunities for students to produce and comprehend language and to obtain modelling and feedback from their peers. Much of the value of cooperative learning lies in the way that teamwork encourages students to engage in such high-level thinking skills as analyzing, explaining, synthesizing and elaborating.

Allen and Van Sickle (1984) did an experimental treatment in a study involving low achieving students. They found that the cooperative learning group scored significantly higher on a world geography test. Perreault (1983) also found that cooperative learning resulted in significantly higher achievement in industrial art students at the knowledge and comprehension levels of Bloom's taxonomy. In another study in which nutrition was taught to both elementary and secondary students using a cooperative learning strategy, Wodarski et al (1980) found significant gains between the pre- and post-test scores. They found that 95% of the elementary students enjoyed the cooperative learning activities and had learned a lot about nutrition. The researchers concluded that cooperative learning was an effective method of teaching nutrition.

After reviewing 46 studies related to cooperative learning, Slavin (1983a) stated that cooperative learning resulted in significant positive effects in 63% of the studies, and only two studies reported higher achievement for the comparison group. Johnson et al (1981) conducted a meta-analysis of 122 studies related to cooperative learning and concluded that there was strong evidence for the superiority of cooperative learning in promoting achievement over competitive and individualistic strategies. Other researches[7] also consistently showed that using experimental/control comparisons of at least four weeks duration, cooperative learning groups demonstrated positive outcomes in measures of achievement, self-esteem, inter-group relations, acceptance of academically handicapped students toward school, and/or ability to work cooperatively.

Yet, there are other studies that seem to indicate the ineffectiveness of group work in certain contexts. For instance, in Ethiopia, a study by Girma (2003) revealed that an overwhelming majority of teachers felt that group work is inappropriate and ineffective to their school contexts. The participants of the study were 74 in-service English language teachers attending special in-service training program at Addis Ababa University. His findings indicated that, in spite of being one of the instructional techniques suggested in the English text books, 79% of the respondents do not implement group work activities. Interview results revealed that because of the large number of students in one class, teachers believed that making students work in group would result in disciplinary problem and uncontrollable noises. As a result, most teachers prefer to avoid group work in order to maintain discipline and silence. Questionnaire data showed that low motivation of students and poor proficiency in English were reported by 71% of the teachers as the second major impediments to the application of group work. Teachers also indicated that since students have very little proficiency in English Language, they do not participate when asked to work in group. What is more, one of the concerns of teachers about implementing group work stemmed from the worry that when students work in group, they are likely to make mistakes and hear each others' English which may not be accurate. Following the findings, Girma recommended that teachers should be provided with training in the use of group work activities in English Language Teaching and

[7] http://www.charlesbridge.com/school/pdf/TAIresearch.pdf (accessed on 29 September, 2005).

acquaint them with the literature on group work. From the review of this research work, however, it is not difficult to observe the lack of strong evidence to support what is reported as the finding (doubting the contribution as well as benefits of group work). Rather the issue seems to be that most of the respondents lack awareness of cooperative learning methods or they seem not to be well equipped with the skills and techniques of applying cooperative learning to enhance student performance. Such factors should have been considered either in the setting of the study or interpretation of the findings.

In summary, the foregoing survey of literature generally revealed that group work has a lot of advantages leading to improved academic performance, attendance and retention, positive relationships among students and a "sense" of community.

2.3 Forming Groups

This section discusses the three issues related to forming groups, namely: (i) determining the size of the group based on the learning objectives; (ii) the allocation of members into groups; and (iii) heterogeneity within groups.

(i) Size of Groups

A number of research works recommended an ideal group size of three to five students (Johnson and Johnson, 1990; Slavin, 1987). A team size of four is especially recommended for a number of reasons[8]. These include:

- During the group work students find it easier to organise meetings as there are less clashes with timetables.
- Students get a larger piece of the work to do and feel they can make a meaningful contribution to the group assignment.
- Students are more visible and accountable to each other. This often reduces the problems associated with the withdrawal of effort.

Apple (2001) also made suggestions on what he called an optimal size of a group based on the learning objectives. His suggestions are presented as follows.

Skill Exercises (teams of two): refer to activities with which students demonstrate their comprehension of new material. Homework problems or questions at the end of a chapter are

[8] http://www.iml.uts.edu.au/learnteach/enhance/groupwork/unit3.html (accessed on August 20, 2005)

examples of skill exercises. The group work is established in such a way that students perform some degree of transfer of knowledge, but contexts are not completely new to students. When the focus of an activity is to strengthen the understanding of previously introduced concepts, working in pairs is desirable. In these situations, it is suggested that the instructor should pair students so that they help each other.

Guided Discovery Learning (teams of three): refers to establishing a group work so that students learn new content through the use of models and a set of guided-inquiry questions. Students learn through discovery rather than being told information directly. However, the mode of discovery is not completely open (as in research) but rather the instructor provides a resource base and guides learners through the process.

In-class Problem Solving (teams of four): allows the instructor to observe and assess problem solving skills of the students. The group work is structured in such a way that students are able to practice problem solving in the presence of peers and an instructor offers feedback to improve future performance.

In the case of in-class problem solving with less complex problems, teams of four are optimal. If the team size is too large for the task at hand, certain team members will be less engaged and contribute less. A team size of four in this situation allows for each team member to make a significant contribution to the process.

Long-term problem solving project (teams of five): This type of activity is typically more of a semester- or term-project in which a team will collectively put in 60 to 100 hours of work. This type of project requires planning at the beginning, incorporation of many concepts from the course and on-going assessment of performance by the team. The final work product most often includes a written report as well as an oral presentation. Teams of five are optimal in this situation. Even with more complex problems, teams greater than five tend to lose members and find a higher degree of disengagement. Also, the team process breaks down more easily with larger teams.

(ii) Allocation of Members into Groups

The following discussions related to methods of allocating members into groups, were obtained from the web site of Interactive Media and Learning[9].

[9] **Institute for Interactive media and learning**
http://www.iml.uts.edu.au/learnteach/enhance/groupwork/unit3.html
(accessed on August 20, 2005)

There are four common methods of allocating members into groups, namely: **Random assignment, Self-selection, Specific criteria and Task appointment**. The first three methods are commonly used when groups are given the same group tasks (for instance, studying in groups, a group essay or report on a pre-defined topic). The fourth assignment method is used when groups are able to choose from a number of pre-set topics. In order to weigh up the options for each, all these four methods are discussed below.

Random Assignment: Many lecturers use some form of random appointment method to form groups. One of the most popular is the call-off system. This is when the lecturer walks around the room and assigns each student in the class a number or letter in a systematic call off (i.e. 1, 2, 3, 4, 5... 1, 2, 3, 4, 5… etc. or A, B, C… A, B, C… etc.). Groups are then formed by putting all the 1's, 2's, etc together. Other random appointment methods include students drawing numbers from a "hat" or the lecturer placing the names of students in the "hat" and then drawing them out.

Random assignment methods are often employed because they are seen as having a number of advantages - they are relatively easy to administer as there is little preparation needed; they break up friendship groups (because most people sit with people they know and they are usually assigned into different groups); they allow people to work with people they ordinarily would not; they are seen by some students as being relatively fair.

However, the random appointment method does have some drawbacks - students feel they do not have any choice in the selection process (particularly those who know others in the class and would prefer to work with them). They worry about the chance of being assigned to a group with incompatible members. They may also consider that the lecturer has used the easiest formation option available. It can send the wrong message to students that the lecturer does not care how the groups are formed.

Self Selection: In many instances, students are asked to form groups by themselves. Under these conditions, students usually know people in their class and choose to work with them. Those who do not know others in the class, tend to form groups with those they are sitting near or with others who may not know anyone either.

The self selection method is easy to administer and students like the opportunity to choose their fellow group members. For many, it is the safe option -- "better the devil you know". However, it can be difficult for students who do not know anyone else in the class and is often seen as not being fair for all students in the class. It often will not yield a desirable level of diversity.

Specific Criteria: This method attempts to form heterogeneous groups. It works on the assumption that groups work better when the members are balanced. Some of the more popular methods use functional roles, learning styles or personalities. These systems involve students completing a questionnaire which is scored to determine preference of a student. Students with different styles are then appointed to each group so as to achieve the desired balance.

There are a number of advantages to the specific criteria method - students feel the method of selection is fair; they see themselves as "experts" and are motivated to demonstrate and apply their skills; they learn about individual differences and how diversity can create synergy.

These methods, while effective, have certain limitations. The group composition process might be expensive and time consuming. The approach may also result in groups with homogeneous characteristics, over-representation of some characteristics or vice versa. This can cause problems when members are assigned to groups (i.e. not all styles or preferences are covered in the group); students may not want to be grouped with students of the same personality traits, skill or preference. For example, they may want to try to develop skills or a style which they do not have.

Task appointment: In this case, the lecturer offers the students a number of topics and lets them select. Groups are then generated from the topics selected. Nomination for the task may involve submitting a preference sheet (students are usually required to rank the topics from most to least preferred) or the students writing their name on a topic sign-up sheet.

The advantages of this approach are that students are more motivated for group work when they choose their own topic; they feel that the selection process of the group is fair; they know they will be working with people who are also interested in the topic and have confidence.

The disadvantages of the approach may be that occasionally, there are too many students wanting to do a particular topic and not enough members selecting others.

(iii) Heterogeneity within Groups

The most widely presented suggestions in the cooperative learning literature are that group composition should be heterogeneous whenever possible. In other words, students in groups should be diverse in background, idea, personality, ethnicity, and gender (Slavin, 1995).

With a desire to innovate and increase student participation, Romney (1996) has employed collaborative learning method to a French translation course in Canada. According to her, the groups, made up of five students, were formed by taking the following factors into account.

- gender (the vast majority of language students were female, so no group contained more than one male);
- language proficiency in both English and French with each group comprising one individual with native or near-native skills in French and one whose first language was neither one of the Canadian official languages;
- personality (for instance, not more than one argumentative or shy student was placed in each team);
- age, work, and life experience;

The resulting groups were as heterogeneous as possible so as to expose students to a variety of opinions. Romney's observation of the groups indicated that, on a personal level, the students were pleased to be able to share their difficulties with others. They gained confidence from observing that if their team-mates could solve problems, they would also be able to overcome them. Speaking in front of a small group with which they were familiar, rather than in front of the whole class, was also less stressful. They also formed close friendships with their team-mates, and many stress that for that reason they look forward to coming to class. Last but not least, on an academic level, there were definite gains in conformity with the findings of Johnson and Johnson (1985) that "cooperative learning experiences promote higher achievement than do competitive and individualistic experiences".

Bradley and Hebert (1997) initiated a study on the effect of individual personality differences on the productivity of a group. Among the preference alternatives in the behaviour of individuals were how a person was energized – designated by extrovert against introvert, The extroverts referred to behaviour of individuals who were energized by interacting with other members of the team as compared with the introverts who prefer to be by themselves. As such, under leadership in a group, the best leader should be an extrovert with either the traits of sensing, thinking and making good judgment or extroverts with keen intuition, deep thinking, and good judgment in making decisions.

In relation to grouping based on ability, Stepaneck (1999) argued that ability grouping was a complex and often divisive issue in education. Heterogeneous grouping is necessary in order to ensure equal opportunities for all students. Students who get stuck in low-level tracks are deprived of opportunities to develop higher-level skills and study rich content (Oakes, 1990). Specialists in ability grouping, made the following recommendations about grouping students:

- Heterogeneous groups are most appropriate when students are working on open-ended problem-solving tasks or science inquiry activities;
- It is also appropriate for students to work in heterogeneous groups when they are discussing concepts that are new to all students;
- Homogeneous groups are more appropriate when students are working on skill development or reviewing material that they have already learned;
- Grouping strategies should be flexible, and students should be allowed to work independently at least occasionally according to their preferences;
- Students should have opportunities to select their own groups based on common interests; and
- All students need to learn the skills of working together before cooperative learning activities will be successful.

As teachers strive to implement collaborative learning strategies and meet the needs of diverse learners, the ability-grouping issue has generated a great deal of research, much of it inconclusive, about the benefits or weaknesses of heterogeneous and homogeneous grouping.

2.4 Cooperative Learning in Mathematics

There are various techniques developed by workers in the field for the purpose of successfully applying cooperative learning methods in the teaching of Mathematics. Following are some examples.

One popular technique is the Jigsaw method first developed by Aronson in the early 1970s Aronson (2000). In the Jigsaw method, students are divided into small groups and are assigned a group task, such as solving a multipart mathematics problem, with several distinct components to it. Each student in a group is then assigned one component of the task, such as a concept or procedure, which she or he must work on independently. After each student completes his or her part, the members of the group reconvene, and each student shares the results of her or his individual work with the rest of the group members. Students are required to listen carefully to each other - otherwise they will not learn all of the material which they will later be individually tested on. After this is done, all of the individual pieces are assembled together to solve the original problem, hence the name Jigsaw.

Panitz (1999) applied the Jigsaw method in the research conducted on cooperative learning in mathematics, particularly in teaching students how to factor polynomials. In the research setting, there were four unique cases in factoring polynomials where the first coefficient is one. Each student in a group is assigned one of those four cases. Through exploration and practice, each student is made to become an expert in his or her assigned case, and then develop a teaching strategy for explaining his or her findings to the other group members. This then resulted in all group members learning how to work with all four cases.

Slavin (1983b) has also researched on what he called Student Teams Achievement Divisions (STAD). The main idea behind STAD was that the scores assigned to teams are based on the extent to which each student improves upon his or her own past record. Students are motivated to actively teach each other the material, to ensure a high group grade. Additionally, since rewards are based upon individual improvement, a group which contains academically poor students is at no disadvantage to a group which contains actively advanced students.

Dees (1983) reported positive results from cooperative learning in mathematics. She noted that when learning something new, students must progress through four types of learning, in this order: facts, skills, concepts, and applications. From her experience, students can best accomplish this when working in a group where they can discuss problems amongst themselves. Additionally, she noted that the cooperative learning method was especially helpful for students who have not previously been successful in mathematics.

2.5 Summary

In our review of literature on performance factors and cooperative learning, attempts were made to present relevant experience and information related to the educational and social psychology component of the current research work. Attributes suggested in the literature as factors to intervene with performance and to bring success in heterogeneous group composition formed the background for the experimental works detailed in chapters 4 to 6.

In relation to cooperative learning, we have seen that such factors as determining the size of a group based on the learning objectives and the grouping methods are important in the group composition process. The focus in this study, particularly, is to research with a method that takes into account both academic performance and personality traits and introduces heterogeneity in the groups. From among the various group composition methods described in the literature, this research focused on the specific criteria method. A group of size four is also considered to be optimal for the purpose of the experiment.

Literature reviewed on Bayesian Networks is presented in the next chapter.

CHAPTER THREE

3. BAYESIAN NETWORKS

In this chapter, the fundamentals of Bayesian probability as well as learning and inference in Bayesian networks are discussed in detail. Some applications of Bayesian networks in the field of education are also presented. The discussions are included as a basis for the performance prediction model in the next chapter. Although complete understanding of the mathematical details may help, they may not as such be necessary to follow the experiment and the results of this study.

3.1 Fundamentals of Bayesian Probability

(i) Basic Axioms

Bayesian probability theory deals with events and the probabilities of those events. If X is an event, then the probability of X is denoted by a real- valued number, P(X). The basic axioms of probability theory (Bayes, 1763; Cowell, 1999) are:

1. P(X) = 1 if and only if X is certainly true.
2. P(X) = 0 if and only if X is certainly false.
3. $0 \leq P(X) \leq 1$ (shows an intermediate degree of belief in the truth of a statement).
4. If X and Y are mutually exclusive, then $P(X \cup Y) = P(X) + P(Y)$.

It is pertinent to define a particular class of events that of a variable A being with certainty in one and only one of the discrete states $a_1 \ldots a_n$. We denote the probability of this event by P ($A=a_i$), and it follows from the axioms that:

$$\sum_{i=1}^{i=n} P(A = a_i) = 1$$

The sequence of probabilities P ($A=a_1$), P ($A=a_2$), …, P($A=a_n$) define a probability vector. A useful shorthand way of referring to this vector is simply P(A).

(ii) Prior and Conditional Probabilities

The unconditional or prior probability associated with a proposition A is the degree of belief accorded to it in the absence of any other information. It is written as P(A). For instance if the prior probability that any student has a high performance in mathematics is 0.3, then we can write "P(performance = high) = 0.3". Prior probabilities are used when there is no other information. As soon as some new information is known, we must reason with the conditional probability of A given this new information .

Accordingly, the notation P(A|B), where A and B are any propositions, is used to represent the conditional probability that A will occur given that B has already occurred. For instance, we might ask "what is the probability that a male student performs high in mathematics?". This might be symbolized as P(performance in mathematics = high | gender=male).

Conditional probabilities can be defined in terms of unconditional probabilities. The defining equation is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{3.1}$$

which holds whenever P(B) > 0;

Or

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{3.2}$$

which holds whenever P(A) > 0

From equation 3.1 and 3.2, we might derive the **product rule**

$$P(A \cap B) = P(A|B) * P(B) = P(B|A) * P(A) \tag{3.3}$$

Similarly, we may define the joint distribution P(A,B,C) as follows:

$$P(A,B,C) = P(A|B,C) * P(B|C) * P(C) \tag{3.4}$$

Rearranging the product rule in equation 3.3 leads to Bayes' famous theorem:

$$\mathbf{P(A|\,B) = \frac{P(B|\,A) * P(A)}{P(B)}} \tag{3.5}$$

Bayes' theorem is frequently used for reasoning about an uncertain hypothesis A given evidence B, and in that context P(A|B) is called the posterior probability of A, and P(A) is called the prior probability of A. The table defining conditional probabilities for every possible combination of values that A and B can take is called a conditional probability table (CPT).

The literature usually shortens $P(A \cap B)$ to P(A,B) and is called a joint probability distribution. Like the conditional probability distribution, it is a table of values, one entry for each possible combination of values that its variables can jointly take. Following equations 3.2 and 3.3 and generalizing for n variables, a joint probability distribution can be defined by the product rule(chain rule) as follows:

$$
\begin{aligned}
P(a_1, a_2,,,,,,,,, a_n) &= P(a_1 | a_2,\ldots, a_n)P(a_2,\ldots, a_n) \\
&= P(a_1 | a_2,\ldots, a_n)P(a_2 | a_3,\ldots, a_n)P(a_3,\ldots, a_n) \\
&= P(a_1 | a_2,\ldots, a_n)P(a_2 | a_3,\ldots, a_n)P(a_3 | a_4,\ldots, a_n)\, P(a_4,\ldots, a_n) \\
&= P(a_1 | a_2,\ldots, a_n)P(a_2 | a_3,\ldots, a_n)\ldots P(a_{n-1}|a_n)P(a_n) \\
&= \prod_{i=1}^{1=n} P(a_i \mid par(a_i)) \qquad (3.6)
\end{aligned}
$$

This property of joint probability distributions is called the general factorization property. Note that this product rule allows for any ordering of variables in the factorization.

While the product rule is used to construct joint probability distributions, marginalization reduces a joint probability distribution to a distribution over a subset of its variables. More specifically, Let a_i denote a state of the variable A. In the table P(A,B), there are m different events for which the variable A is in **state** a_i, (namely the mutually exclusive events $(a_i, b_1), \ldots, (a_i, b_m)$). Therefore, $P(a_i)$ can be calculated as;

$$P(a_i) = \sum_{j=1}^{j=m} P(a_i, b_j) \qquad (3.7a)$$

For example, suppose we have the joint probability distribution P(A,B), as shown in the following conditional probability table(CPT).

Table 3.1: An example of a joint probability distribution

	b_1	b_2	b_3
a_1	p_{11}	p_{12}	p_{13}
a_2	p_{21}	p_{22}	p_{23}
a_3	p_{31}	p_{32}	p_{33}
a_4	p_{41}	p_{42}	p_{43}

$P(a_1)$ may then be computed as

$$P(a_1) = \sum_{j=1}^{j=3} P(a_1, b_j) = P(a_1,b_1) + P(a_1,b_2) + P(a_1,b_3)$$

$$= p_{11} + p_{12} + p_{13}$$

In this operation, we say that the variable B is marginalized out of P(A,B) (producing P(A)). Thus, the notation can be written as follows:

$$P(A) = \sum P(A, B) \tag{3.7b}$$

(iii) Conditional Independencies

Dependence between two events is when the probability of an event depends on the knowledge of the other event. Suppose that the factor P(A|B,C) has the property that it is always equal to P(A|C). That is, for every pair (a,c), P(A=a|B,C=c) remains constant as B varies. We, therefore, say that A is conditionally independent of B given C. We can therefore drop B from the conditional probability P(A|B,C) altogether and rewrite the representation as:

$$P(A|B,C) = P(A|C).$$

Consequently, we can also write:

$$P(A,B,C) = P(A|C)P(B|C)P(C) \tag{3.8}$$

The use of conditional probabilities in probabilistic reasoning is analogous to the use of hypothetical assumptions in logical reasoning. However, conditional probability has the advantage that conclusions can be stated more flexibly as probabilities, rather than as true/false statements.

This completes our brief introduction of the basic concepts of probability theory. The remainder of this chapter will deal with Bayesian networks.

3.2 Basics of Bayesian networks

Bayesian networks are graphical representations of causal relations in a domain. They are composed of a set of variables, a set of directed links between variables and with each variable is associated a conditional probability table(Jensen, 1996). For instance, two variables (also referred to as “nodes”) A and B are connected by directed link from A to B, if A has a causal impact on B. The intuitive meaning of a link in a properly constructed network is, A has a direct influence on B. In this particular case, we find the conditional probability table P(B|A) attached with node B.

In a directed graph, the terminology of family relations is adopted to explain the relations between the variables. If there exists a directed link from variable A to variable B, then A is called a parent of B and B is called a child of A. The variables symbolize events. Every variable in a Bayesian network has two or more discrete[10] states (i.e. colour of a car: blue, green, red, etc, gender: male or female, performance of a student: high, low, or average).

We now present a more formal definition of Bayesian networks.

Pearl (1988) represented a Bayesian network by $BN = \langle N, A, \theta \rangle$, where $\langle N, A \rangle$ is a directed acyclic graph (DAG)[11] – each node $n \in N$ represents a domain variable and each arc $a \in A$ between nodes represents a probabilistic dependency between the associated nodes. Associated with each node $n_i \in N$ is a conditional probability table (CPT), collectively represented by $\theta = \{\theta_i\}$, which quantifies how much a node depends on its parents (Pearl, 1988).

Russel and Norvig (2003) stated that the structure of a Bayesian network is a graphical illustration of the interactions among the set of variables that it models. They explained the full specification as follows:

(i) A set of random variables makes up the nodes of the network;

(ii) A set of directed links connects pairs of nodes. If there is a directed link from node A to node B, A is said to be a parent of B;

(iii) Each set contains a finite set of mutually exclusive states;

[10] In general, variables can have continuous or discrete states.

[11] Directed Acyclic Graph(DAG) refers to a graph where all of the edges in the graph are directed and there are no cycles. (There is no way to start from any node and travel along a set of directed edges in the correct direction and arrive back at the starting node).

(iv) Each node A has a conditional probability table P(A| par(A)) that quantifies the effect of the parents on the node. If the variable A does not have any parent, then the table can be replaced by prior probabilities, i.e. P(A);

(v) The variables coupled with the directed edges construct a directed acyclic graph(DAG).

In Bayesian networks, it is reasonable to suppose that each random variable is directly influenced by at most K others for some constant K (Russell and Norvig, 2003). If we assume "n" Boolean variables for simplicity, then the amount of information needed to specify each conditional probability table will be at most 2^k numbers and the complete network can be specified by $n2^k$ numbers.

An example of a Bayesian network is shown in Figure 3.1. With every node is associated a table of conditional probabilities of the vertex given the state of its parents. We denote the conditional probability table using the notation $P(x_i \mid par(x_i))$, where lower case x_i denotes values of the corresponding random variable X_i and $par(x_i)$ denotes a state of the parents of X_i. The graph together with the conditional probability tables defines the joint probability distribution contained in the data.

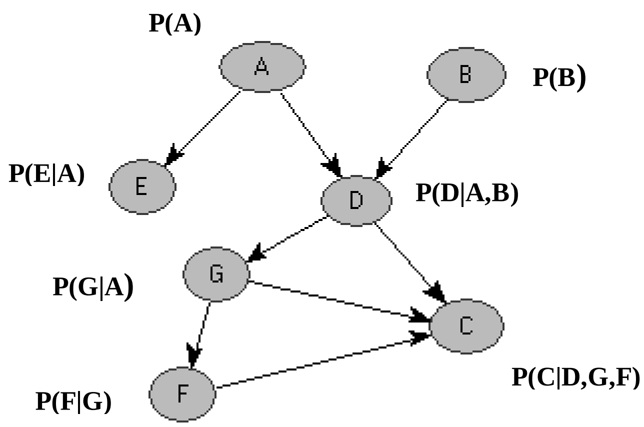


Figure 3.1: A Bayesian Network (the probabilities to specify are shown)

Using the probabilistic chain rule, (based on equation 3.6), the joint distribution can be written in the product form. For instance:

$$\mathbf{P(C,F,G,D,E,A,B)} = P(C|F,G,D,E,A,B)*P(F|G,D,E,A,B)*P(G|D,E,A,B)*P(D|E,A,B)*P(E|A,B)*P(A|B)*P(B)$$
$$= P(C|F,G,D)*P(F|G)*P(G)*P(D|A,B)*P(E|A)*P(A)*P(B)$$

The main advantage of Bayesian networks is the ability to define the conditional independencies first, before specifying numerically the actual conditional probability distributions. A general conditional independence property of Bayesian networks is that any variable X in the network is conditionally independent of its non-descendents ND(X) given its parents par(X) (Pearl, 1988). That is, if a variable's parents become known, then any information about nodes that are not on a directed path from X will be irrelevant. This is the so-called directed Markov property of Bayesian networks.

Figure 3.2 is specific example of a Bayesian network extracted from the performance prediction model. The four variables considered are: gender (G), interest for mathematics (I), shyness (S) and mathematics performance (M). A particular instance of the Markov condition is also shown, where mathematics performance(M) is independent of its non-descendants shyness(S) and gender(G) given its parent - interest for mathematics(I).

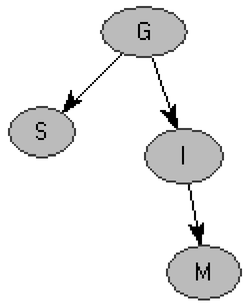


Figure 3.2: Example of Bayesian network consisting of four attributes.

A direct arc between G and I denotes the fact that whether or not a student is male or female will influence the likelihood of high interest for mathematics. Similarly an arc from G to S denotes that gender influences the extent of shyness. Arc I→M means that interest for mathematics affects the level of performance in mathematics.

Lack of directed arcs is also a way of expressing knowledge which asserts conditional independence. For instance, the absence of arc S→I means that a student's shyness does not affect interest for mathematics. The absence of arc I→S has a similar interpretation. These causal assertions can be translated into statements of conditional independence. Mathematics performance is conditionally independent of shyness and gender given interest. In mathematical notation:

$$P(M|I) = P(M|I,S) = P(M|I,G) = P(M|S,I,G)$$

The causal knowledge also translates into probabilistic assertions: S and I are conditionally independent given G.

$$P(S,I|G) = P(I|G) * P(S|G)$$

or

$P(S|I,G) = P(S|G)$, Which is another instance of Markov condition.

These properties imply that

$$P(M,S,I,G) = P(M|S,I,G) * P(S|I,G) * P(I|G) * P(G)$$
$$= P(M|I,G) * P(S|G) * P(I|G) * P(G)$$

In relation to graphical illustration of Bayesian networks, there are generally three types of connections, namely Serial, Diverging and Converging connections. They are depicted in Figure 3.3(a), 3.3(b) and 3.3(c).

(i) **Serial:**

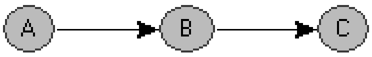


Figure 3.3(a): A serial connection

In **serial connections**, A has control over B which then has control over C. Apparently, the evidence[12] on the variable A will affect the certainty of the variable B that in turn affects the certainty of variable C. Analogously, the evidence on the variable *C* will affect the certainty of the variable *A* through the variable *B*. On the contrary, if the state of the variable *B* is given, then the link is blocked, and variables A and C become independent. Influence can pass from A to C and vice versa unless B is instantiated. In other words, if B becomes known, then it effectively separates A and C. In other words, a path between nodes A and C is closed, given some evidence B, if A and C are conditionally independent given B (Pearl, 1988). A and C are then said to be D-separated (Direction separated). A more detailed explanation of D-separation will be given later.

(ii) **Diverging:**

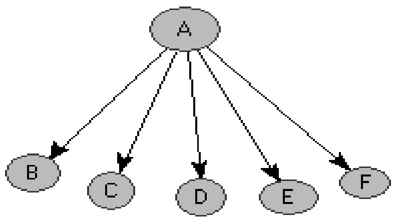


Figure3.3(b): A diverging connection

As shown in Figure 3.3 (b), in a **diverging connection** the influence can pass between all the children of the variable *A* unless the state of the variable *A* is given. If the state of the variable

[12] Jensen (1996) stated that evidence on a variable is a statement of the probabilities of its states. If the statement supports the exact state of the variable, it is called hard evidence. Otherwise, it is called soft evidence.

A is known, then the variables *B*, *C*, D, E and F become independent from each other. Therefore, influence may run between *A*'s children unless *A* is instantiated.

(iii) **Converging:**

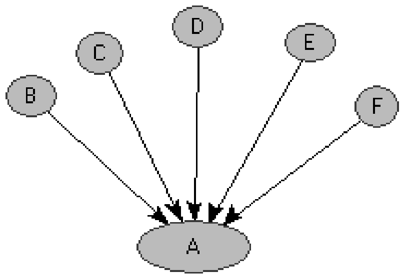


Figure 3.3(c): A converging connection

In a **converging connection** such as the one shown in Figure 3.3 (c), if there is nothing known about the variable *A* other than what may be deduced from the knowledge of its parents *B*, *C*, D, E, and F, then the parents are said to be independent. The independence implies that evidence on one of the parents has no effect on the certainty of the others. If there is any other kind of evidence influencing the variable *A*, then the parents become dependent because of the principle of explaining away[13]. Therefore, evidence may only be transmitted through a converging connection if either the variable in the connection or one of its descendants has received evidence. The evidence can be direct evidence on the variable *A*, or it can be evidence from one of its children.

The three connections explained above wrap all the forms in which evidence may be transmitted though a variable. It is observed that one can decide for any pair of variables in a causal network whether or not they are dependent once knowing the evidence entered into the network.

Before we conclude the introduction, the following Bayesian network semantics are worth mentioning. By applying Bayes' theorem, the direction of the arcs can be reversed as long as a

[13] **Explaining away** is the process of decreasing one's belief in a causal event as a result of an increase in the belief of an alternative causal event. For example, in the above converging network, suppose B is the proposition that Student X has mastered the topic and C is the proposition that Student X performs poorly on exams. If A is the proposition that Student X failed the exam, then B and C are two possible explanations for A and would therefore converge to a causal network. If we subsequently observe A (say, to find that the proposition is true and the student did fail the exam), then A's causes become dependent because if B is subsequently observed to be true as well, then it (intuitively) has some bearing on C (meaning the student X failing the exam has been explained by performing poorly on exams).

directed cycle is not induced. While changing the arc, directionality may change the d-separation properties of the network, the overall joint probability distribution will be invariant. Therefore, technically, networks differing only in arc directionality can be considered equivalent. However, semantics are conventionally used to make particular configurations of arc directions unique. While not entailed by the underlying theories, the addition of semantics is convenient. The most common interpretation of an arc is: if A is a parent of B, then A is said to exert a causal influence on B, or precede B temporally, and not the other way around (Jensen, 1996).

3.3 Learning in Bayesian Networks

3.3.1 Introduction

A Bayesian network may be hand-constructed by a domain expert, i.e., the domain expert draws the dependencies between the nodes. The conditional probabilities can then be assessed by the expert, learned from data, or obtained using a combination of both techniques (Neapolitan, 2004). However, eliciting Bayesian networks from experts can be a laborious and difficult procedure in the case of large networks. In order to address the problem, researchers have developed methods that could learn the DAG (structure learning) as well as the conditional probability distributions from data (parameter learning).

The process of learning Bayesian networks takes different forms in terms of whether the structure of the network is known and whether the data is complete. We actually have four cases of learning Bayesian networks from data;

(i) Unknown network structure and complete data,
(ii) Known network structure and complete data,
(iii) Unknown network structure and incomplete data, and
(iv) Known network structure and incomplete data.

Learning with complete data indicates that the training data contains no missing values, while incomplete data indicates that some piece of information in the data are not known.

To date, there exist a variety of Bayesian network learning algorithms for each of these situations (Eg. Cooper and Herskovits, 1992; Spirtes et al, 1993).

Learning Bayesian networks with known network structure and complete data is the most studied case in the literature, since the network structure is already defined and the algorithm needs to estimate only the parameters (Spiegelhalter et al, 1993). Parameter learning is achieved simply by calculating the conditional probability table(CPT) entries using estimation techniques such as Maximum Likelihood Estimation and Bayesian estimation. An approach to parameter learning with complete data is described by Heckerman (1999) and Krause(1999).

In cases where the network structure is unknown and the data is complete, the learning algorithm is given the set of variables in the model and needs to select the arcs between them and to estimate the parameters. This problem is especially useful with no available domain expert and when we want to get all of the benefits of a Bayesian network model. In addition, the learnt structure also gives the expert some indications of what attributes are correlated.

In general, the Bayesian learning method, with unknown structure and complete data, involves three phases (Druzdzel and Diez, 2003). These phases are,

(i) Manual selection of the model variables and their possible values;

(ii) Automatic determination of the structure of the graph based on a Dataset (i.e., learning what depends on what). There are, basically, two main approaches for learning the structure of a network from complete data:

- **Constraint-Based**: perform tests of conditional independence on the data, and search for a network that is consistent with the observed dependencies and independencies, according to the concept of d-separation (Pearl 1988). Identification of conditional independence relationships among the variables is done using some statistical test (such as chi-squared test). One can find the conditional independence relationships among the attributes and use these relationships as constraints to construct a BN (Spirtes et al., 1993; Cheng et al., 1997).

- **Score-based**: define a score that evaluates how well the (in)dependencies in a structure match the data, and search for a structure that maximizes the score. This suggests that the best BN is the one that best fits the data (Heckerman, 1995; Cooper and Herskovits, 1992).

(iii) Automatic calculation of the conditional probability distribution over each of the model variables (i.e., strength of dependencies by entries in CPT).

In order to explain how a Bayesian network is developed from complete data, (both in the case when a network structure is known and when it is not) we present one of the publicly available learning algorithms namely the **Three Phase Dependency Analysis(TPDA)** learning algorithm. The TPDA is a constraint based learning algorithm (Cheng and Greiner, 2001).

In a preliminary investigation made for the purpose of selecting an appropriate learning algorithm from those which are publicly available, this algorithm was found to perform better and hence its selection for this research work.

3.3.2 The Three Phase Dependency Analysis (TPDA) Learning Algorithm

The goal of the TPDA learning algorithm is to find what is connected to what – i.e., which nodes should be joined by arcs. The algorithm works incrementally: at each point, it has a current set of arcs, and is considering adding some new arc, or deleting an existing one. Such decisions are based on information flow between a pair of nodes, relative to the rest of the current network.

In this algorithm, a Bayesian network is viewed as a network of information channels or pipelines, where each node is a valve that is either active or inactive and the valves are connected by noisy information channels(arcs). Information can flow through an active valve but not an inactive one. Suppose two nodes X and Y are not directly connected within a network structure – if the structure is correct, then there should be no information flow between these nodes after closing all of the existing indirect connections between X and Y. The learning algorithm, therefore, tries to close off all of these connections, then asks if the dataset exhibits additional information flow between these two nodes. If so, the learner will realize the current structure is not correct, and will add a new arc (pipeline) between X and Y.

The TPDA algorithm is viewed as a constraint based learning algorithm, since it uses conditional independence test results as constraints. Therefore, domain knowledge can be incorporated in a natural way as constraints. In the remainder of this section, definitions and basic concepts used in TPDA are presented followed by a discussion of the algorithm in relation to both unknown and known structure.

(i) Basic Concepts in TPDA

In order to render a meaningful discussion, the following concepts and definitions are supported by illustrations from the following simple multi-connected graph (Figure 3.4).

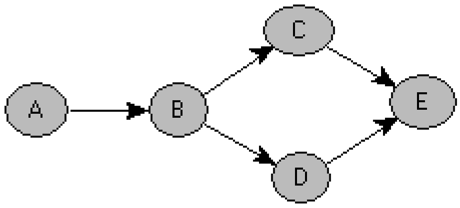


Figure 3.4: A simple multi-connected Bayesian Network

Adjacency Path

For two nodes X and Y, an "Adjacency path"[14] $P = \langle a_1, a_2, a_3 .....a_k \rangle$ between a_1 = X and a_k = Y is a sequence of arcs that if viewed as undirected edges, would connect X and Y. In Figure 3.4 above, C-E-D is an adjacency path connecting C and D, even though the arcs are in different directions.

Collider

A node V is a collider of the path $\langle a_1, a_2, ...a_{i-1} = (X,V), a_i = (Y,V), ..., a_k \rangle$ if the two directed arcs associated with that node, here $a_{i-1} = (X,V)$ and $a_i = (Y,V)$, collide at V. In other words, if two arcs in the path meet at their end point on node V, we call V a collider of the path. A node that is not a collider of a path is called a non-collider of the path. The concept of a collider is always related to a particular path. In Figure 3.4 above, we say that E is a collider in the path C-E-D.

Conditional Independence(CI)

Let X,Y,Z be any three variables. X and Y are said to be conditionally independent given Z if for all events x ∈ X, y ∈ Y and z ∈ Z P(x|y,z)=P(x|z) whenever P(y,z)>0.

D - Separation

For a DAG G=(N,A), for any nodes $X, Y \in N$ where $X \neq Y$, and "evidence" $L \subseteq N \setminus \{X, Y\}$, we say that "X and Y are d-separated given L in G" if and only if there exists no open adjacency path between X and Y, where any such adjacency path P is considered **open** if and only if

[14] To distinguish it from the directed path that connects two nodes by the arcs of a single direction, we call this kind of paths adjacency paths or chains.

(a) every collider on P is in L or has a descendent in L and

(b) no other nodes on path P is in L.

In Figure 3.4, given empty evidence, C and D are d-separated. Putting a node into the cut-set is equivalent to altering the status of the corresponding valves – hence, putting the collider E into the cut-set will open the path between C and D; while putting a non-collider B into the cut-set will close both the A-B-C-E and the A-B-D-E paths, thereby d-separating A and E.

Dependency Map, Independency Map and Perfect Map

A graph G is a **dependency map (D-map)** of a probabilistic distribution P if every dependence relationship derived from G is true in P; G is an **independency map (I-map)** of P if every independence relationship derived from G is true in P. If G is both D-map and I-map of P, we call it a **perfect map (P-map)** of P, and call P a DAG-Isomorph of G (Pearl, 1988). Here we say that P and G are faithful to each other (Spirtes et al., 1993).

Node Ordering

Node ordering is a kind of domain knowledge used by many Bayesian network learning algorithm that specifies a causal or temporal order of the nodes of the graph (variables of the domain). For instance, in Figure 3.4, node B can not happen earlier than node A.

Mutual Information

The volume of information flow between two nodes A and B is measured using mutual information, which is defined as follows:

$$I(A,B)=\sum P(a,b)\log\frac{P(a,b)}{P(a)P(b)} \tag{3.9}$$

And the conditional mutual information, with respect to the set of "evidence" nodes C is given by

$$I(A,B\mid C)=\sum P(a,b\mid c)\log\frac{P(a,b\mid c)}{P(a\mid c)P(b\mid c)} \tag{3.10}$$

This mutual information between A and B measures the expected information gained about B, after observing the values of the variable A. In Bayesian networks, if two nodes are dependent, knowing the value of one node will give us some information about the value of the other node. Hence, the mutual information between two nodes can tell us if the two nodes are dependent and if so, how close their relationship is.

Given the actual probability distribution P(x), we would claim that A and B are independent if I(a,b) = 0. However, the learning algorithms do not have access to the true distribution P(x) (total population) but instead use empirical estimates, based on a Dataset D. The learning algorithm therefore uses $I_D(A,B)$ which approximates $I\ (A,B)$ and uses $P_D(X)$ rather than P(X). The algorithm, therefore, claims that A is independent of B whenever $I_D(A,B) < \varepsilon$ for some suitable small threshold[15] ε > 0. Similarly, conditional independence is declared whenever $I_D(A,B \mid C) < \varepsilon$.

As described by Cheng et al. (1997), the TPDA algorithm makes the assumption that the higher the mutual information between two variables in the data, the more likely it is that an arc should connect them in a Bayesian network.

(ii) TPDA with Unknown Structure and Complete Data

In what follows, the TPDA algorithm for learning without node ordering(unknown structure) is presented. The algorithm takes a database table as input and constructs a Bayesian network structure as output. Since node ordering is not given as input, this algorithm has to deal with two major problems (i) how to determine if two nodes are conditionally independent and (ii) how to orient the edges in a learned graph.

As described by Cheng et al (1997), the algorithm has four phases: ***drafting***, ***thickening, thinning*** *and* ***orienting edges***. The steps involved in each of these phases are explained below.

Phase I: (Drafting)

1. Initiate a graph G(V,E), where V={all the attributes of a data set}, E ={}, Initiate an empty list L.
2. For each pair of nodes (v_i, v_j) where v_i, v_j element of V and i different from j, compute mutual information $I(v_i, v_j)$ using equation 3.9.

   For all the pairs of nodes that have mutual information greater than a certain small value ε, sort them based on their mutual information values and put these pairs of nodes into list L from large to small. Create a pointer **p** that points to the 1st pair of nodes in L.

[15] In the description of the dataset used to test the algorithms the value of the threshold ε was set to 0.01

3. Get the 1st two pair of nodes of list L and remove them from L. Add the corresponding arcs to E. Move the pointer **p** to the next pair of nodes.
4. Get the pair of nodes from L pointed to by the pointer **p**. If there is no open path between the two nodes, add the corresponding arc to E and remove this pair of nodes from L.
5. Move the pointer **p** to the next pair of nodes and go back to step 4 unless **p** is pointing to the end of L.

To illustrate the algorithm in action, consider the induction of a Bayesian network with five nodes(variables) *A,…, E*. Suppose eight pairs of nodes have a mutual information greater than $\mathcal{E}$, and that they are ordered from highest mutual information to lowest yielding *L* = <*A-B, B-E, E-C, A-C, B-C, A-D, D-C,D-E*>. Now, <*A-B*, *B-E*, and *E-C*> are added directly to the network since there is no path already between them. *A-C* and *B-C* cannot be added because the addition of the first three edges already resulted in a paths connecting *A* and *C*, and *B* and *C*. *A-D* is next on the list to be added, and after this is done, the network has become singly-connected. The edges remaining in *L* are <*A-C, B-C, D-C, and D-E*>. The output of this phase is one singly-connected network spanning the entire network (i.e. there is at most one chain between any two nodes, as shown in Figure 3.5).

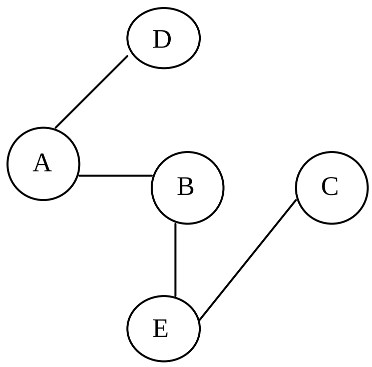


Figure 3.5: The network after the drafting stage

Phase II: (Thickening)

6. Move the pointer P to the first pair of nodes in L.
7. Get the pair of nodes (node1, node2) from L at the position of the pointer **p**. Call procedure find-cut-set[16] (current graph, node1, node2) to find a cut-set that can d-separate node1 and node2 in the current graph. Use a conditional independence test to

[16] The TPDA algorithm uses **find-cut-set** procedure to get a cut-set that can d-separate the two nodes, and then uses a CI test to see if the two nodes are independent conditional on the cut-set. This procedure tries to find a minimum cut-set (a cut-set with minimum number of nodes). After the CI test, an arc is added if the two nodes are not conditionally independent.

see if node1 and node2 are conditionally independent given the cut-set. If so, go to the next step; otherwise, connect the pair of nodes by adding a corresponding arc to E.

8. Move the pointer **p** to the next pair of nodes and go back to step 7 unless **p** is pointing to the end of L.

After this stage, the algorithm is guaranteed to have found all the edges in the final Bayesian network. In our example, the graph after phase II is shown in Figure 3.6. Arc A-C is added because A and C are not independent conditional on ⟨B⟩, which is the smallest cut-set between A and C in the current graph. Arc B-C and D-C are added for similar reasons.

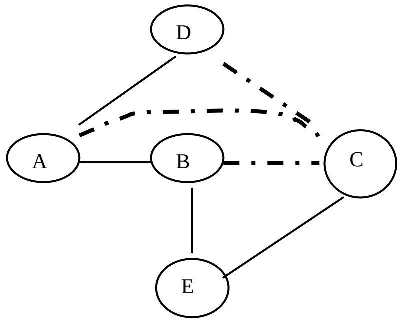


Figure 3.6: The network after the thickening stage (New edges added from *L* are dashed)

However, unwanted surplus edges may have been added as a result of the linear order in which edges are added to the network from *L*. The task of phase III is to identify those wrongly added arcs and remove them.

Phase III (Thinning)

9. For each arc(node1, node2) in E, if there are other paths besides this arc between the two nodes, remove this arc from E temporarily and call procedure find-cut-set(current graph, node1, node 2) to find a cut-set that can d-separate node1 and node2 in the current graph. Use a conditional independence test to see if node1 and node2 are conditionally independent given the cut-set. If so remove the arc permanently; otherwise add this arc back to E.

The output of this step is the final (undirected) structure of the Bayesian network.

In our example above, note that edge *A*-*C* was added to the network before edges *B*-*C* and *D*-*C*. If it is the case that the addition of these latter two edges results in a cut-set (e.g. {*B*,*D*}) d-separating *A* and *C*, then the thinning step would remove *A*-*C* permanently.

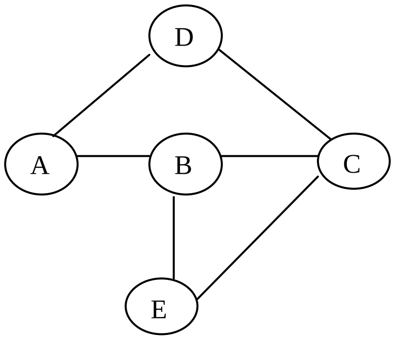


Figure 3.7: The network after the thinning stage (Note that the edge *A*-*C* has been dropped).

Phase IV (Orienting Edges)

10. The final phase orients the edges with the concept of collider, converging and divergent networks.

Consider the node *B* on a path *A*-*B*-*C* in Figure 3.7. If *B* is a converging connection, then *B*'s neighbours *A* and *C* on the path will be independent until *B* or one of its descendents is instantiated, at which point they become dependent. Therefore we can analyse the data to determine all the triplets of variables along a path in the network having this property, and thereby identify all the converging connections in the network. The remaining nodes must be either serial or diverging connections. When an edge cannot be oriented, the orientation task is left to the domain expert. In the case of our example, nodes *D* and *B* are found to be converging connections on all their paths, which allows every edge except *E*-*C* to be orientated.

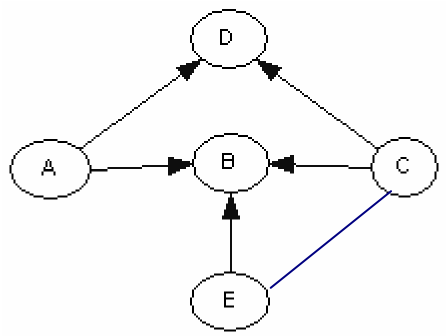


Figure 3.8: The network after its edges have been oriented. (Note that edge *E*-*C* cannot be oriented).

The algorithm based on the above four stages is shown below.

Subroutine TPDA (D:Dataset, $\mathcal{E}$:threshold): returns G=(V,E): graph structure

Begin [Drafting]

1. Let V={attributes in D}, E={ }
   L ={<X,Y>|I(X,Y)> $\mathcal{E}$} be the list of all pairs of distance nodes <X,Y> where X,Y $\in$ V and X $\neq$ Y, with at least $\mathcal{E}$ mutual information

2 Sort L into decreasing order, wrt(I(X,Y)

3. For each ⟨X,Y⟩ in L:
   If there is no adjacency path between X and Y in current graph (V,E)
   Add ⟨X,Y⟩ to E and
   Remove⟨X,Y⟩ from L.

Begin [Thickening]

4. For each ⟨X,Y⟩ in L:
   If EdgeNeeded_H((V,E), X,Y:D, $\mathcal{E}$)
   Add ⟨X,Y⟩ to E

Begin[Thinning].

5. For each ⟨X,Y⟩ in E:
   If there are other paths, besides this arc, connecting X and Y,
   E'= E - ⟨X,Y⟩ %---- temporarily remove this edge from E
   If $\neg$EdgeNeeded_H(V,E'), X,Y; D, $\mathcal{E}$ % i.e, if X can be separated from Y.
   % in current "reduced" graph.
   E = E' % then remove <X,Y> from E

6. For each ⟨X,Y⟩in E:
   If X has at least three neighbors other than Y, or Y has at least three neighbors other than X,
   E' = E - <X,Y> % i.e., temporarily remove this edge from E
   If $\neg$EdgeNeeded((V,E'), X,Y;D, $\mathcal{E}$
   %i.e. If X can be separated from Y in current "reduced" graph
   E = E' % then remove ⟨X,Y⟩ from E

7. Return [OrientEdges((V,E),D)]

Figure 3.9: TPDA Algorithm without node ordering (Cheng et al, 1997)

(iii) TPDA with Known Network Structure and Complete Data

We now briefly describe the TPDA-II learning algorithm where node ordering is given, i.e., the algorithm takes as input both a table of database entries and a node ordering and constructs a Bayesian network structure as output.

The first three phases of this algorithm are the same as the TPDA algorithm described in the previous section. However, the last phase(orienting edges) described above, is not implemented in this algorithm, since the direction of the arcs are decided by the node ordering provided. The main features involved in these three phases are

(i) When direct cause and effect relations are available, it uses them as a basis for generating a draft in phase I.

(ii) In phase II, the algorithm will try to add an arc only if it agrees with the domain knowledge.

(iii) In phase III, the algorithm will not try to remove an arc if it is already specified by domain experts.

The TPDA-II algorithm is shown in Figure 3.10 below

```
Subroutine TPDA-II (D:Dataset, π : node ordering, ε: threshold):
Returns G = (V,A): graph structure

  1.    Let V: ={attributes in D}, A:={}
        L:={(X,Y) | I(X,Y) >ε} be the list of all pairs of distinct nodes(X,Y)
        Where X,Y∈ V and X ≠ Y in π , with at least ε mutual information.
                    Begin [Thickening]
  2.    For each ⟨X, Y⟩ in L:
        C:= MinCutSet(X,Y; (V,A), π )
        If I_D(X,Y | C) > ε
            Add (X,Y) to A
                    Begin [Thinning]
  3.    For each (X,Y in A:
        If there are other paths, besides this arc, connecting X and Y,
            A':= A – (X,Y)         % i.e., temporarily remove this edge from A.
            C := MinCutSet(X,Y; (V,A'), π )
            If I_D(X,Y | C) < ε
                    % i.e., if X can be separated from Y in current "reduced" graph
            A:=A'          % then remove ⟨X, Y⟩ from A
      4.    Return (V,A)
```

Figure 3.10: The TPDA-II algorithm – node ordering given (Cheng et al, 1997)

3.4 Inference in Bayesian Networks

3.4.1 Introduction

Bayesian inference is the general problem of computing the posterior probability $P(\mathbf{Q}|\mathbf{E}=\mathbf{e})$, for some evidence **E=e** and query **Q** (where $\mathbf{Q} \subseteq \mathbf{X}$ and $\mathbf{E} \subseteq \mathbf{X}$). It is fairly simple when it involves only two related variables. However, it becomes much more complex when we want to do inference with many related variables, particularly using the manual approach. For instance, we may want to do probabilistic inference involving features that are not related via a direct influence.

To illustrate how the manual approach is intractable and inefficient, consider an example of a Bayesian network depicted in Figure 3.11. The Bayesian network depicts the performance prediction model in which student's motivation(M) affects confidence (C) which in turn affects the extent of shyness(S). Student's extent of shyness affects attitude towards group work(A). The student's extent of shyness affects English performance(EP) which in turn affects mathematics performance(MP). Shyness(S) and interest for mathematics(I) are affected by student's gender(G). With this Bayesian network, we can perform inferential queries such as P(C|M=low), or P(MP| S=Introvert, EP=Satisfactory, G=male).

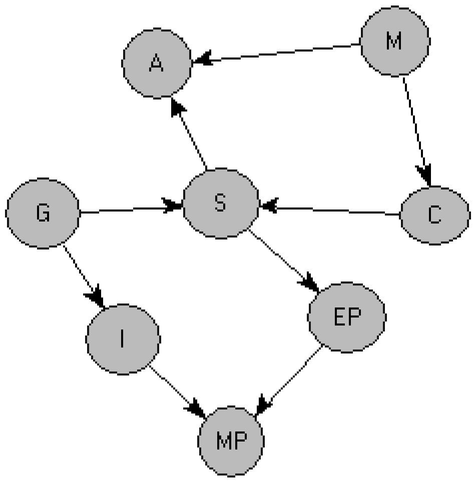


Figure 3.11: Example Bayesian network

P(MP,A,EP,I,S,G,C,M) =

P(MP|EP,I,)*P(A|M,S)*P(EP|S)*P(I|G)*P(S|G,C)*P(G)*P(C|M)*P(M)

(from eqn. 3.6)

This approach is in general difficult, time consuming and generally error prone. For even the simplest query, the marginalization steps require summations and multiplications over all the variables in the network.

To tackle this problem of computation, a variety of Bayesian network inference algorithms have been investigated. These inference algorithms can be roughly classified as **exact** or **approximate.** Among the several approximate algorithms developed based on stochastic sampling, the best known are probabilistic logic sampling (Henrion, 1988), likelihood sampling (Shachter and Peot 1990; Fung and Chang 1990), and backward sampling (Fung and del Favero 1994).

There also exist several efficient exact inference algorithms, that make belief updating in graphs consisting of tens or hundreds of variables tractable. Pearl(1986) developed a message-passing scheme that updates the probability distributions for each node in a Bayesian network, in response to observations of one or more variables. Lauritzen and Spiegelhalter (1988) and Jensen et al. (1990) proposed an efficient algorithm that first transforms a Bayesian network into a tree where each node in the tree corresponds to a subset of variables in the original graph. The algorithm then exploits several mathematical properties of this tree to perform probabilistic inference.

The most commonly used algorithm for BN inference is the clique tree algorithm (Jensen et al, 1990; Lauritzen and Spiegelhalter,1988). The inference algorithm developed by Lauritzen and Spiegelhalter(1988) and later clarified by Jensen(1990), is used in this research work for predicting level of performance of a student. The remainder of this section, therefore, presents the concepts behind this algorithm.

3.4.2 The Lauritzen/Spiegelhalter Algorithm

The basic approach of the Lauritzen Spiegelhalter algorithm(hereafter referred to as LS algorithm) is to transform the Bayesian network into a singly-connected structure, and then perform local computations on that structure rather than the original network. This algorithm basically has two procedures: **compilation** (graphical and numerical) procedure, in which the input is the original Bayesian network specification and the output is the singly-connected structure, and a **propagation** procedure, in which evidence is absorbed and queries are performed on the new structure.

(i) Graphical Compilation

The **graphical compilation** procedure involves taking the original Bayesian network and transforming it into a *junction tree*. The junction tree representation is equivalent to the original Bayesian network, except that it is singly-connected even if the original network was multiply-connected. The generation of a junction tree requires five steps, namely: **marry co-parents, moralise network, triangulate network, form junction graph of cliques, form junction tree**.

The **marrying** of co-parents is the simple addition of an arc between any two nodes that are parents of the same child, but not already neighbours. The **moralisation** of the network is the dropping of all arc directions. The output of these two steps, when applied to the example Bayesian network in Figures 3.11, is shown in Figure 3.12. The non-adjacent co-parents in the original network are (EP,I), and (G,C), and so edges between these pairs are added to the network. (The new edges are dashed in Figure 3.12).

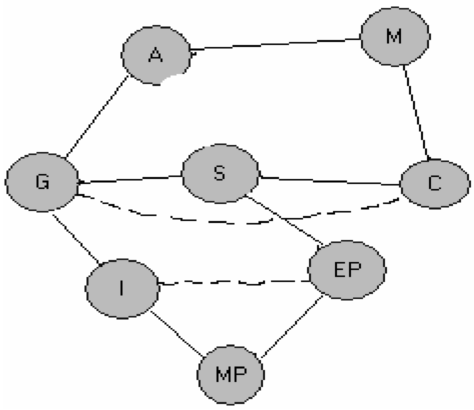


Figure 3.12: Graph after the marry, moralize step

In the **triangulation step**, arcs are successively added to every cycle of length 4 or more that does not already have an arc, until no such cycles exist. To illustrate, consider Figure 3.12. A number of cycles exist in this graph, such as G-S-C and EP-MP-I. However, neither of these are candidates for shortening because they are cycles of length 3 (not 4 or more). There are only two cycles in Figure 3.12 appropriate for shortening. They are the cycles S-EP-I-G and G-A-M-C-S. In order to shorten these cycles, new edges S-I , A-C and A-S are included that make the graph fully triangulated. The result is depicted in Figure 3.13.

Note that graph triangulation may not always be necessary. It is quite possible that following the marrying and moralisation steps of the compilation, the network will already be triangulated and therefore, the execution of the triangulation algorithm may be unnecessary.

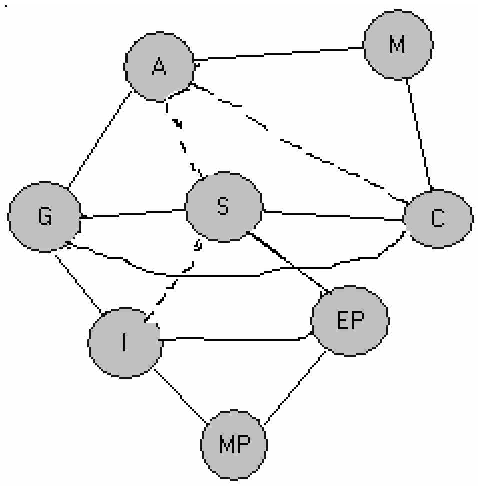


Figure 3.13: The triangulated graph

The fourth step involves **identifying the cliques in the triangulated graph**, and forming a new graph called a junction graph. A clique is a "maximal, complete" subgraph where every node in the subgraph is adjacent to every other node. For example, ⟨EP,MP,I⟩ is a maximally complete subgraph in Figure 3.13 since there is no other node that can be included in this subgraph. ⟨EP,S,C⟩ is not because there is no edge EP-C. ⟨S,G⟩ is complete but not maximally so, because ⟨S,G,C⟩, the subgraph formed by adding C, is complete. The cliques of Figure 3.13, therefore, are: ⟨A,C,M⟩, ⟨A,C,S⟩, ⟨**A**,S,G⟩, ⟨G,S,C⟩, ⟨G,S,I⟩, ⟨I,S,EP⟩, ⟨I,EP,MP⟩. In the junction graph, each node corresponds to a clique. Since there are seven cliques in Figure 3.13, there will be seven nodes in the junction graph.

Furthermore, variables from the original graph are likely to appear in more than one clique; to capture this in the junction graph, an edge is added between two cliques if their intersection is non-empty. Figure 3.14 is the junction graph derived from Figure 3.13.

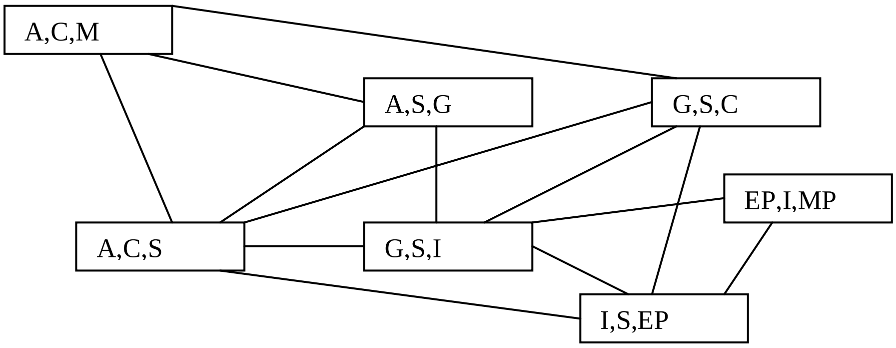


Figure 3.14: The junction graph

As mentioned earlier, the motivation for the compilation stage is to produce a singly-connected structure so that inference via local computation is possible. This structure, the **junction tree**, is formed by simply "pruning" the junction graph until only a tree remains. However, the junction tree has an additional property not present in the junction graph; namely, the *running intersection property*: if any two cliques in the junction tree contain a mutual variable *X* from the original network, then every clique on the path between those two cliques must also contain *X*. This ensures that the junction tree does not have two or more disconnected "representations" of the same variable. The running intersection property thus restricts the way in which a junction graph can be "pruned" to a junction tree. A junction tree for Figure 3.14, labelled with the clique intersections, is depicted in Figure 3.15.

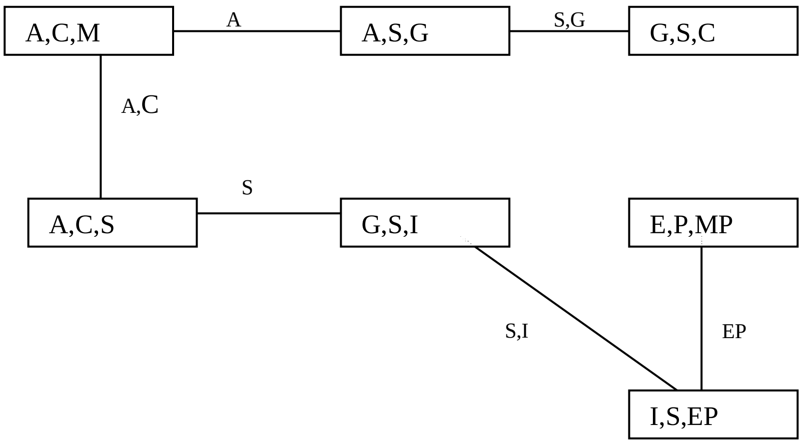


Figure 3.15: A junction tree with the running intersection property

Once the junction tree is constructed, the numerical compilation procedure is performed in order to make a logical mapping between the original form of the Bayesian network and its recursive factorisation[17]. More mathematical details of this process are available in Lauritzen and Spiegelhalter (1988) and Jensen et al (1990).

(ii) Propagation

Once the junction tree is constructed and the representation of the Bayesian network is transformed into a product of clique marginals using the numerical compilation procedure, evidence can be propagated and queries performed. Graphical Propagation on a junction tree starts with a single clique receiving evidence, and its neighbours successively *calibrate* themselves to absorb the evidence. The evidence "flows" via the variables that are the intersection of the neighbouring cliques.

The numerical propagation part of the LS algorithm is basically the propagation of consistency from a clique to its neighbours in a junction tree. Consistency is a property belonging to pairs of neighbouring cliques, and is achieved when marginalising on the variables shared by both neighbours yields the same belief distribution. More mathematical details of this process are available in Lauritzen and Spiegelhalter (1988) and Jensen et al. (1990).

3.5 Applications of Bayesian networks

During the past decade, Bayesian networks have gained popularity in Artificial Intelligence as a means of representing and reasoning with uncertain knowledge. They are increasingly being used in expert systems for diagnosis, forecasting, decision analysis, control theory application and intelligent agent modelling. The goal in using them is to capture dependencies that exist in real decision–making problems.

Some domain specific works have focused on probabilistic student models. The Andes Intelligent Tutoring System for Physics (Vanlehn et al., 2002), uses a belief network to represent alternate plans that may be used to solve physics problems. Student actions are analyzed to update the probabilities of the respective plans. Conati and VanLehn (1996) and Vanlehn and Martin (1995) presented an On-Line assessment of Expertise (OLAE) that collects data from student solving problems in introductory college physics and analyzes the data with probabilistic methods that determine what knowledge the student is using and presents the results of the analysis. For each problem, the system automatically creates a

[17] With the final junction tree, for each variable x, there is one and only one factor P(X|par(X)) in the recursive factorization.

Bayesian net that relates knowledge represented as first-order rules, to particular actions, such as written questions. Using the resulting Bayesian network, OLAE observes the behaviour of a student and computes the probabilities of the level of knowledge of the student and accurate use of rules.

The research presented by Murray (1998) inferred a student model from performance data using a Bayesian belief network. The belief network modelled the relationship between knowledge and performance for either test items or task actions. The measure of how well a student knows a skill is represented as a probability distribution over skill levels. Questions or expected actions are classified according to the same categories by the expected difficulty of answering them correctly or selecting the correct action.

In summary, Bayesian networks have been found to have the following advantages:

- They handle incomplete data sets without difficulty because they discover dependencies among all variables;
- One can also learn about causal relationships between variables using Bayesian networks and the strength of the causal relationships with probabilities
- Considering the Bayesian statistical techniques, Bayesian networks facilitate the combination of domain knowledge and data. Prior or domain knowledge is crucially important if one performs a real-world analysis; in particular, when data is inadequate or expensive. The encoding of causal prior knowledge is straightforward because Bayesian networks have causal semantics;
- Independencies can be dealt with explicitly. They can be articulated by an expert, displayed graphically, and reasoned about, yet they remain robust to numerical expressions.
- Bayesian Network structure represents the inter-relationships among the attributes. Humans can easily understand the network structures and experts can modify them to obtain a better predictive model.

CHAPTER FOUR

4. IDENTIFICATION AND MEASUREMENT OF ATTRIBUTES

This chapter presents the survey conducted in order to finalize the selection of relevant attributes that apply to the local context, from among the attributes identified in the literature and discussions with experts. The survey was designed in such a way that common attributes that intervene with mathematics performance and relevant to bringing heterogeneity into groups could be identified. The identification of common attributes was considered in order to reduce the complexity of the work.

In the first section, the setting of the experiment (i.e., issues applicable in the whole experiment) is briefly outlined. This is followed by sections dedicated for attribute selection, the measurement process, data organization/analysis and data preparation for the experimentation.

4.1 Setting of the Experiment

4.1.1 Test Targets

As indicated in the introductory chapter, the research targeted high school students who are in the preparatory program (for College/University education). Information obtained from the Ministry of Education showed that, there were 22 public and 35 private high schools in Addis Ababa which ran preparatory programs. Of these, Yekatit 12 Senior Secondary School was selected. This is a famous public school located near the main campus of Addis Ababa University. The track records of the school showed that every year (for the last couple of decades) a good number of its graduates join universities as compared to students of other public high schools. In relation to the student population, information obtained from the record office of the school indicated that as of the 2003/2004 academic year, there were 1,215 students in grade 12, where 608 were in the morning shift and 607 in the afternoon shift.

4.1.2 Data Protection and Privacy Issues

Prior to the conduct of all surveys, attempts were made to address the data protection and privacy issues by explaining the main objectives of the study to the school administrators, teachers and students. The researcher was also introduced as a Ph.D. student with an official support letter from the Dean of the Faculty of Informatics, Addis Ababa University. Because of the sensitive nature of the data to be collected which also involved the identification of each individual student who filled the data gathering instrument, students were asked, by their mathematics instructors, for their consent to participate in the experiment. In addition, they were assured that the data supplied/collected would remain confidential and be destroyed once the experiment was completed. Oral instructions as well as written general and specific directions were also given to the students to emphasize honesty in their responses. On a personal note, during the different phases of the experiment which took almost more than a year, both the students and instructors were very friendly and did actively participate in the study.

4.1.3 Subject Area

As indicated in Chapter 2, of the various group learning objectives (skill exercises, guided discovery learning, in-class problem solving and long-term problem solving projects), the case of in-class problem solving type is considered for the purpose of experimentation. Moreover, in order to contextualize the research work, mathematics was selected as the subject area. The factors considered in picking mathematics for the purpose of this study include the following.

- Familiarity of the researcher with the subject (teaching freshman mathematics for more than five years);
- The relationship between mathematics performance and academic or Career opportunities. In most institutions, a successful performance in mathematics is used as one of the selection criteria both for placement in higher education and employment (Mills, 1993);
- Because of its vital importance in the school curriculum, education systems throughout the world place high importance on the teaching and learning of mathematics and a lot of effort is being made to improve efficiency and effectiveness in these activities (Garden, 1987).

- Despite the importance mentioned, many students seem to have wrong impressions about mathematics and dislike mathematical activities; many seem to fear, even hate mathematics (Neale, 1969). As a result, mathematics is becoming unpopular as a subject (Banks, 1964, Ernest, 1976).

At the initiation of this research project and to get more insight into the local situation, two preliminary studies were conducted in February, 2002: one on review of letter grades of four batches of freshman students (1998-2001) and the other a survey on the reflections by students on learning Freshman Mathematics at the College of Social Sciences.

As observed in the preliminary surveys (Rahel, 2002), a greater share of freshman students score low grades in mathematics. In general, 50% of the freshman students out of the four batches got a letter grade of "C" or less in mathematics. The students have found the mathematics course difficult regardless of the fact that a higher proportion of the topics they learned were revisions from their high school mathematics. Some of the difficulty experienced include: difficulty in keeping pace while the instructor is teaching, lack of provision for adequate and well organized tutorials to help them practice more and poor instructional methods resulting in a decline in motivation of learners.

In view of the foregoing, it was felt appropriate to focus on mathematics for the purpose of satisfying the data requirements of this research and consequently develop a resource that would, in the long term, contribute to popularize this important subject among students.

4.2 Selection of Attributes

As indicated earlier, performance prediction and group composition require the identification and study of relevant attributes. The candidate attributes produced from the literature were further reviewed (additional items included) in consultation with domain experts. The whole exercise identified 14 attributes (See Appendix A).

The list of attributes was validated by senior instructors with relevant and rich experience in teaching, research, group composition as well as measurement and evaluation techniques in the field of education. These instructors were selected from Departments of Educational Psychology, Mathematics, Sociology, Information Science and Foreign Languages at Addis Ababa University. The selection process identified about 30 instructors and all of them volunteered to participate in the validation process.

The instructors were asked to indicate the extent of their agreement to consider the attribute in group formation and whether the same attributes would also be a factor for determining level of performance. The responses were put in 3 scales: **Agree**, **Undecided** and **Disagree**.

Once the responses of the instructors were collected, those attributes that were considered as a factor for both performance and group composition were picked and tallied based on frequency count. The list was then organized in frequency order and ranked as shown in Table 4.1 below.

Table 4.1: Rank order of attributes considered

Attributes	**Ranking**
Mathematics Performance	1
Interest for Mathematics	1
English language ability (English performance)	2
Achievement motivation	3
Seriousness/Dedication	4
Group Work Attitude	5
Gender	6
Self confidence (Internal Locus of Control)	6
Age	7
Shyness (introvert personality)	8
Religion	9
Ethnic background	10
Family educational background	11
Family economic background	12

Based on a close examination of the responses and a further detailed discussion with the respondents, the following issues were raised and addressed.

- **Achievement motivation** and **seriousness** are strongly related, and they were combined and renamed as “Achievement Motivation”.
- Considering the target group, it was observed that most students in the preparatory program were in a similar age group. Therefore, regardless of **age** being ranked seven, it was not necessary to include it as an important attribute.
- **Religion** and **Ethnic background** were considered to be sensitive areas when taking into account the current economic and political conditions of the country, and thus were excluded from further consideration.

- In consultation with experts, and in due consideration of the profiles of students in public schools in Ethiopia who come from families that belong to more or less the same category, it was decided to exclude **family educational and economic background** from further consideration as it would not make significant difference among the target population. Besides, the data obtained from students on economic background of parents may not be reliable.

Based on the aforementioned, the attributes identified as factors for performance were **Interest for mathematics, English performance, Achievement motivation, Group work attitude, Gender, Self confidence, and Shyness**. Moreover, these attributes together with the predicted Mathematics performance were used in the group formation process (Details are provided in Chapter 6).

The following operational definitions of the attributes were considered for the purpose of the study.

- **Gender** referred to the sex of the student.
- **Group work Attitude** referred to the way a student viewed and tended to behave towards group work. It was used to explain the feelings of a student about group work particularly associated with studying/learning in groups in and out of school.
- **Interest for mathematics** referred to the liking/disliking the student developed towards mathematics.
- **Achievement motivation** referred to the disposition of a student to approach success. It was used to explain the activated state of a student to get a high standard in his academic performance.
- **Self Confidence** referred to the belief of a student in himself/herself or the student's internal/external locus of control.
- **Shyness (Introvert Personality)** referred to the feeling of being insecure when the student was among other people or talking with other people.
- **Performance in English:** referred to the performance level of a student in English exams. (It also referred to the ability of the student to properly read, write and understand the English language).
- **Performance in Mathematics**: referred to the level of performance of a student in mathematics tests.

4.3 Attribute Measurement Process

The next step after the identification of the attributes was the design of ways to obtain values for these attributes. This actually formed the basis for obtaining the experimental data and building the performance prediction model based on Bayesian networks. While the values of the attributes **English performance** and **Mathematics performance** were obtained from student school records, the values of the other attributes, namely, **Gender, Group work attitude**, **Interest for mathematics**, **Achievement motivation**, **Self confidence** and **Shyness** were obtained based on a data collection instrument designed for the purpose.

This section presents the details of the design and development of the instrument employed for the purpose of collecting the experimental data. Subsequent sections are detailed descriptions of the data collection process.

(i) Development of Instruments

The instrument was first developed by collecting a pool of items to measure each attribute. A number of existing instruments were consulted. Expert opinions were also solicited including the experiences of other researchers for the purpose of determining how to measure each attribute. A total of 28 items to measure group work attitude, 25 items to measure interest for mathematics, 30 items to measure achievement motivation, 28 items to measure self confidence and 29 items to measure shyness were developed. While some items were written in a positive (pro) direction, others were written in a negative (con) direction.

The items for each of the personality attributes were further customized, modified and rewritten to provide better measurement scheme appropriate to the cultural and social conditions in Ethiopian. This was considered relevant because of the fact that cultural assumptions that are widely accepted in Ethiopia about the nature of family relationships, teacher student relationships and the interpersonal dynamics of relationships between students may not be shared by other cultures. Each item in the instrument was assessed for cultural appropriateness based on discussions with experts in the field. The items were first prepared in English and, to further equalize the language understanding level, they were translated to Amharic (a language spoken and understood by almost all students). The instrument was also designed in such a way that there was no special cognitive level required to read and understand the items.

After putting both the English statements and the Amharic translations together, they were distributed to some volunteer instructors in the Departments of Psychology, English, Statistics, Mathematics, Curriculum and Instruction and Amharic. This was done mainly to validate the soundness of the translations made.

While checking the correctness of the translation, each instructor was further asked to give his/her opinion on which attribute each of the group of items were measuring. Opinions given were more or less similar in context in the case of group work attitude, interest for mathematics, and self confidence. Achievement motivation and shyness needed further revision, since the opinions were quite different from what the items were intended to measure. The translated versions were, thus, modified based on the comments of the respondents.

In the next step, 25 items were selected and carefully rewritten for each of the personality attributes. The researcher made use of 10 inter-raters (professional judges): seven raters from the Department of Educational Psychology, one from the Department of Mathematics and two from the Department of Information Science. Information on the qualifications and other related data of the judges who participated in the rating of the items is attached as Appendix B. The Judges were asked to indicate those items best suited for measurement of the personality attributes (i.e. does the item measure what it intended to measure?), and also to indicate the items which not at all measure the variable). In order to create a common understanding between the raters, the description of each attribute to be measured was attached together with the items to be rated.

The procedure employed in evaluating the inter-judge agreement was to count the frequency of the agreement on each item. Items which were agreed by at least eight raters (more than 75%) to measure the attributes were considered for the initial pilot test. Accordingly, as shown in Table 4.2, 21 items to measure attitude, 20 items to measure interest for mathematics, 17 items to measure achievement motivation, 20 items to measure self confidence, and 20 items to measure shyness were retained for the initial pilot test. Items indicated as ambiguous by some of the raters were either further modified or taken out.

The following table shows the personality attributes and the number of items developed for each variable in the first pilot test.

Table 4.2: Number of items developed for each attribute- first pilot test

Personality attribute	Number of items	Positively worded	Negatively worded
Group work Attitude	21	18	3
Interest for mathematics	20	13	7
Achievement motivation	17	14	3
Self Confidence	20	14	6
Shyness(introvert personality)	20	4	16

The rules for assigning numbers in measurement were set according to Likert scale[18]. As a result of discussions made with experts, and for the purpose of convenience, the terms "**Strongly Agree** " and "**Agree"** were categorized into one as **"Strongly Agree"** and similarly the terms "**Strongly Disagree"** and "**Disagree"** were categorized into one as **"Strongly Disagree"**. The Likert scale was, therefore, minimized into three scales. A pilot test was then conducted in order to validate whether the instrument was properly formatted for the intended users and suited to provide with accurate measurement of attributes.

(ii) Item Analysis

The pilot test was tried out in Yekatit 12 Senior Secondary school with one section of students in the morning shift. After the explanation on the data protection and privacy issues, a total of 64 students were made to fill out the test items. After completion, they were asked to give feed back on difficulty level and clarity of the items.

The responses of the students for each item were scored based on the score values of each item. In the case of positively worded items, a score of 3 (high value) was given for a "strong agreement", a score of 2 (average value) was given for "agreement to some extent" and a score of 1 (low value) was given for "strong disagreement". In the case of negatively worded items, a score of 3 (high value) was given for a "strong disagreement", a score of 2 (average value) was given for "agreement to some extent" and a score of 1 (low value) was given for "strong agreement". The data entry was then made using the SPSS statistical package. This package has in-built functions to carry out the necessary statistical tests and examine the reliability of the items in measuring the corresponding attribute.

[18] Likert scale is a five point scale in which the interval between each point on the scale is assumed to be equal.

Statistical tests were carried out in order to find the correlation[19] of each item with total score, a case of construct validity (Guilford, 1956). After computing the item-total correlation coefficient for each item, they were ranked in order of magnitude of their correlation. Based on discussions with domain experts, those items with item-total correlation less than 0.20 were discarded. After removing those items, the Cronbach alpha[20] (index of reliability) was computed for the remaining items measuring each attribute. This was carried out in order to ensure the internal consistency of the items developed and to confirm that the instrument developed elicited a consistent and reliable measure for the attributes. The Cronbach coefficient computed for each of the attributes is given in the table below.

Table 4.3: Cronbach alpha computed for each of the attributes – first pilot test

Attribute	**Alpha**
Group work attitude	0.5686
Interest for Mathematics	.6563
Achievement Motivation	.3057
Self Confidence (Internal Locus of Control)	.4713
Shyness(Introvert personality)	.4944

Because of the low value of alpha, attempts were made to rephrase some items and new items were also included. A second pilot test was, therefore, found necessary before making the final survey. Those items that were marked as not clear by the students were also modified in the second pilot test. For the purpose of validation, the lie detector statements[21] were also included in the second pilot test.

[19] An item-total coefficient of correlation indicates the item's discrimination value--that is, whether or not the scores on the item differentiate between those persons who score high and those who score low on the test as a whole. In general the value should be greater than 0.20.

[20] **Cronbach alpha**, also referred to as an **Index of Reliability**, is a numerical coefficient assessing reliability of scales measuring an attribute. It measures how well a set of items measures an attribute. The coefficient ranges in value from 0 to 1 and may be used to describe the reliability of items measuring the attribute in question. The higher the score, the more reliable the generated scale is. Nunnaly (1978) has indicated 0.7 to be an acceptable reliability coefficient. See Cronbach (1951, 1970) for further reference.

[21]**Lie detector statements** refer to statements which are exaggerated where, in a normal environment, agreement is extremely unlikely except for careless responses or outright lying in order to appear respectable or competent.

The second pilot test was conducted in the same school using another section of students in the morning shift. A total of 60 students participated in the second pilot test. Table 4.4 shows the number of items developed for each variable in the second pilot test.

Table 4.4: Number of items developed for each attribute - second pilot test

Personality attribute	Number of items	Positively worded	Negatively worded
Group work attitude	15	12	3
Interest for mathematics	15	12	3
Achievement motivation	15	13	2
Self Confidence	15	11	4
Shyness	15	7	8
Lie Detector statements	8	8	0

After entering the data into the SPSS package, assessment of the items in the second pilot test indicated that they had a higher item-total correlation as compared to the 1st pilot test. Some of the items having an item-total correlation less than 0.2, particularly to measure achievement motivation and shyness, were rephrased before the final administration of the instrument. The following table presents the Cronbach alpha for the items measuring each attribute.

Table 4.5: Cronbach alpha computed for each of the attributes – second pilot test

Attribute	**Alpha**
Group work attitude	.8322
Interest for Mathematics	.9393
Achievement Motivation	.7909
Self Confidence (Internal Locus of Control)	.7862
Shyness(Introvert personality)	.7696

The final instrument prepared (attached as Appendix C) was a four page instrument consisting of two parts. While the first part consisted of demographic characteristics, the second one measured the different attributes discussed above.

(iii) Test Administration

Sample Size

From the pilot test and from discussion with experts in the field of statistics, the determination of the sample size depended on the chance of happening of a rare event. For example, getting a high value for some of the attributes was considered to be a rare event. From the pilot test, the probability of getting a student with high achievement motivation was the least (0.0625). Therefore, this probability was used to calculate the minimum sample size required which, according to Cochran(1977), could be computed as follows.

$$n = \frac{t^2 pq}{d^2}$$

Where n = the sample size; t = 1.96 for alpha = 0.05; p = the probability value of the rare event(0.0625); q = 1-p = 0.9375 and d = .02

By substituting the values in the above formula, the minimum size of the samples to be taken was set to 562. This was almost 50% of the size of the total population. (i.e., with a two shift system being practiced in the school this sample size means considering all students in one shift).

The margin of error 'd' which is 2% actually resulted with a sample size of 562. With this sample size, one expected to find minimum 4.25% of the students to be highly motivated and maximum 8.25% (i.e. 6.25 ±2.0).

Procedure of Data Collection

There were 10 sections in the afternoon shift of grade 12, each having an average of 60 students. These students were all considered in order to meet the minimum sample size required. The researcher, with the help of two colleagues[22], administered the instrument to the students and a total of 571 instruments were collected. The average time spent by the students while completing the instrument was about 30 minutes.

In addition to the use of the instrument, English and Mathematics results of the students for three consecutive semesters, were obtained from the school records.

[22] These colleagues were given orientation on how they would administer the instrument.

4.4 Data Organization and Analysis

For the purpose of preparing the learning data for the Bayesian network learning algorithm as well as to make it suitable for the group composition process, the collected data was further organized and analyzed as described below.

i) Editing

Before the actual data entry, the lie detector statements inserted in the instrument were used to check for consistency and seriousness. Those instruments where more than 50% of the lie detector statements were filled in as “Strongly Agree” were discarded. Some instruments which were not properly filled were also taken out. A total of 514 returned instruments were retained for the final data analysis.

(ii) Data Entry and Organization

The scores used for each item were the same as those used during the pilot test. The SPSS statistical package was used both for entry and organization of the collected data. Table 4.6 is a summarized information of the background of students who filled out the instrument.

Table 4.6: Background data on student Samples

Category	Classification	Number of Respondents
Sex	Male	301
	Female	213
Age	16	29
	17	188
	18	208
	≥19	89
Educational background of Fathers	Elementary	167
	Secondary	138
	Diploma	89
	First Degree	82
	Second Degree and above	29
	Missing[23]	9
Educational background of Mothers	Elementary	244
	Secondary	139
	Diploma	103
	First Degree	14
	Second Degree and above	6
	missing[23]	8

[23] Missing are students who did not specify their respective demography.

(iii) Data Analysis

After entering the scored items into the SPSS package, they were summed up for each of the attributes. i.e., for each student, the total score on each personality attribute was calculated - the maximum being 45 and the minimum being 15. With regard to the English and Mathematics marks, they were converted to their corresponding Z-scores in order to control for teachers' teaching and grading differences. The standard deviation was then calculated by combining all the 514 students. A sample of the resulting values is shown in Table 4.7.

Table 4.7: Sample table showing values of attributes for each data record.

Gender	Group Work Attitude	Interest for Math	Achievement Motivation	Self Confidence	Shyness	English Performance	Math Performance
Male	43	33	33	36	29	1.18	-1.60
Male	40	30	32	34	22	-0.55	-1.88
Male	34	39	32	32	28	-1.07	1.55
Male	37	37	41	44	38	1.4	2.37
Male	36	39	24	33	29	-2.26	-0.74
Male	31	31	31	33	25	-0.94	-2.24
Male	28	34	35	32	25	2.05	6.48
Male	41	28	35	33	28	-3.6	-2.01
Male	34	38	35	37	37	-0.19	1.69
Male	42	43	39	38	37	3	2.20
Male	32	41	38	37	42	-0.69	-2.32
Male	25	28	34	31	21	-2.05	-2.24
Female	44	43	41	39	42	2.36	4.11
Male	38	32	33	40	35	-1.45	-2.59
Female	36	33	30	32	19	-2.89	-1.08
Male	32	37	28	29	25	3.37	4.66
Male	36	29	33	37	22	-3.18	-0.44
Female	39	43	36	38	33	-0.64	2.61
Male	37	41	34	33	24	2.81	4.21
Male	34	30	35	33	29	-3.9	-0.82
Male	36	39	39	37	33	-2.23	-2.17
Female	38	29	34	35	25	1.39	-2.56
Male	43	42	39	40	25	-2.9	0.01
Male	40	38	39	37	33	-0.81	-3.48
Male	40	25	31	40	34	3.18	-2.60
Female	41	34	41	36	29	-4.1	-2.64
Male	32	39	35	40	29	-3.35	-1.40
Female	40	33	32	37	31	-2.29	-2.87
Female	44	40	40	29	24	1.1	-0.56
Male	38	39	32	36	34	4	0.45
Female	40	44	44	38	35	0.09	1.34

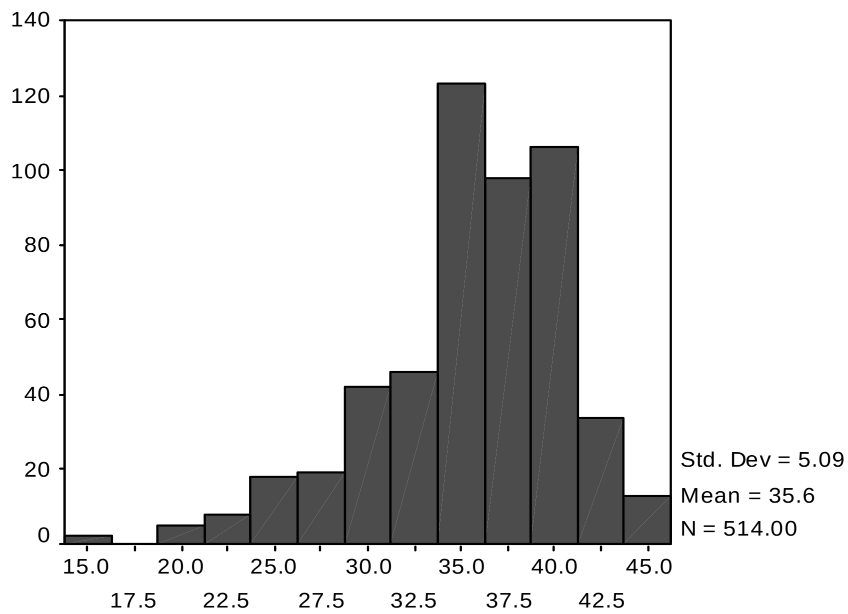
The bulk of the effort was invested in preparing input for belief network investigation. Typographical errors in the data were avoided because each value of the attribute was an SPSS generated one. In addition, data entry errors in the values were detected by graphing each of the attributes.

The observations plotted for each personality attribute were found to be approximately normally distributed (as shown in the graphs below) with the following mean and standard deviation.

Table 4.8: Mean and standard deviation of measured attributes

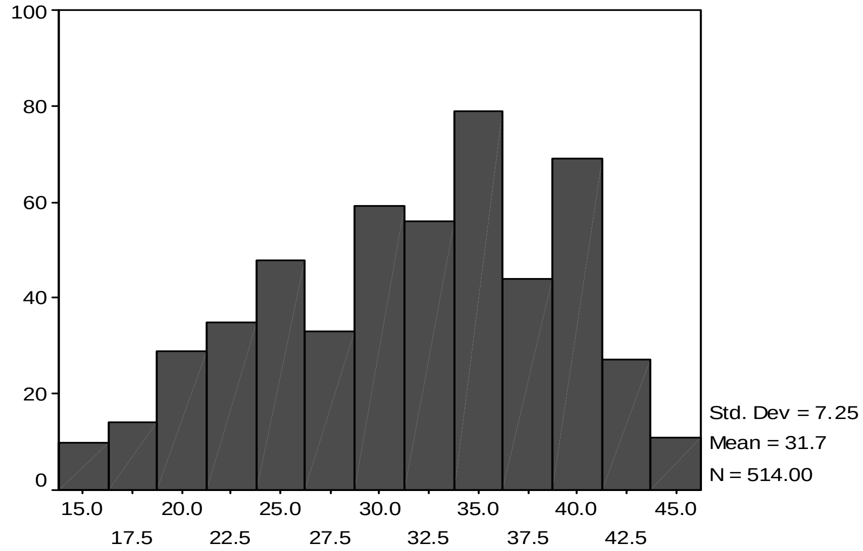
Attribute	**Mean**	**standard deviation**
Group work attitude	35.6	5.09
Interest for Mathematics	31.7	7.25
Achievement Motivation	34.9	4.70
Self Confidence	35.8	3.58
Shyness	28.9	6.11
English Performance	0	2.45
Mathematics Performance	0	2.47

The distributions generated by the SPSS package for the five attributes are shown below.

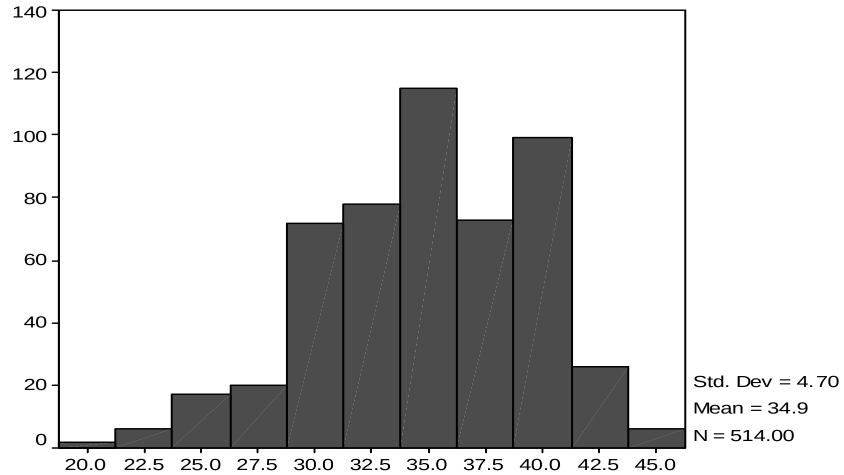


Graph 4.1: Distribution of values for group work attitude

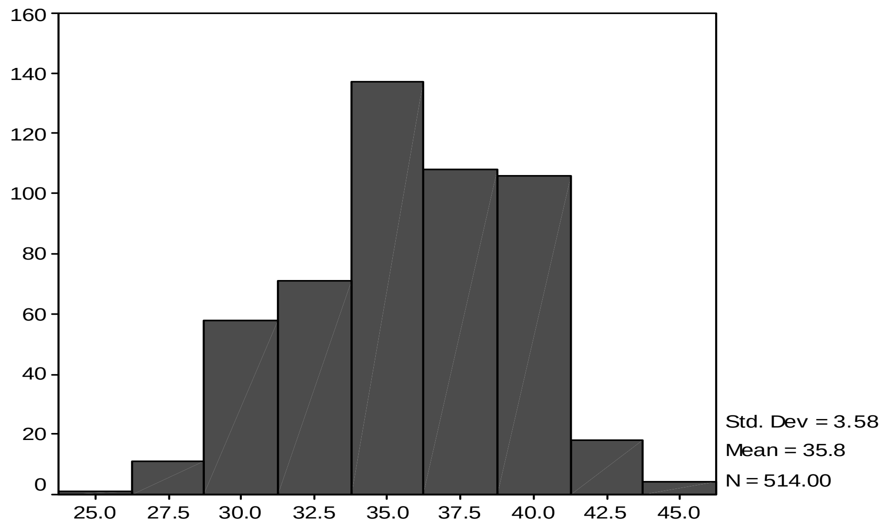
Note: For all the distributions, X-axis refers to sum of the scores and Y-axis refers to the number of students)



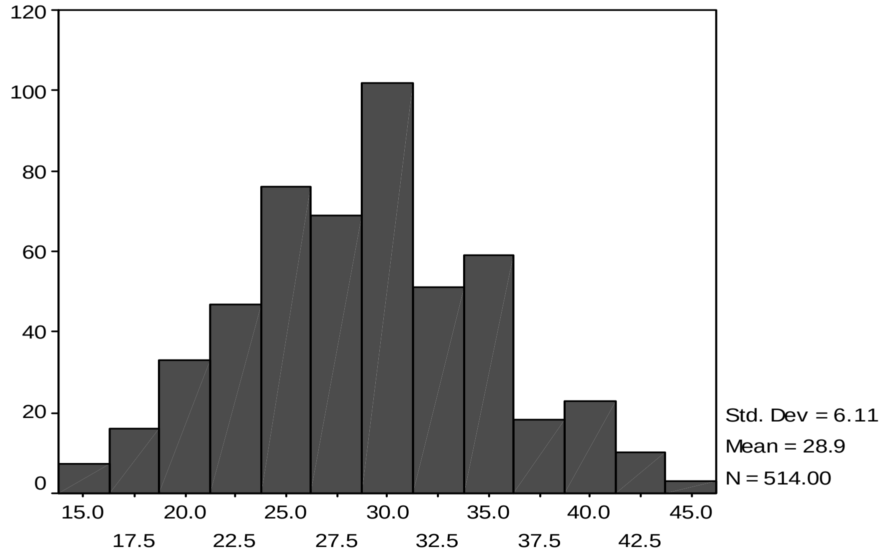
Graph 4.2: Distribution of values for interest for mathematics



Graph 4.3: Distribution of values for achievement motivation



Graph 4.4: Distribution of values for self confidence



Graph 4.5: Distribution of values for shyness

Assuming normal distribution, 66% of the observations were expected to lie within one standard deviation of the mean. This was used to categorize the attributes into three values. As depicted in Table 4.9, observations above the sum of mean and standard deviation were grouped in the high-value category. Those below the difference of mean and standard deviation were grouped in the low-value category and those in between were grouped in the average-value category.

Table 4.9: Categories of attributes

Attributes	Interval (X±1S)	High-value category	Average-value category	Low-value category
Group work attitude	(30.51,40.69)	≥ 40.69	30.51 . . .40.69	≤ 30.51
Interest for Mathematics	(24.45,38.95)	≥ 38.95	24.45 . . .38.95	≤ 24.45
Achievement Motivation	(30.2, 39.6)	≥ 39.60	30.20 . . .39.60	≤ 30.2
Self Confidence	(32.22, 39.38)	≥ 39.38	32.22 . . . 39.38	≤ 32.22
Shyness	(22.79,35.01)	≥ 35.01	22.79 . . . 35.01	≤ 22.79

For English and mathematics performance, the individuals whose total z score was above the standard deviation were considered to have above satisfactory performance, those between minus one standard deviation and one standard deviation were considered to have satisfactory performance and those below minus the standard deviation were considered to have below satisfactory performance. Table 4.10 summarizes the category values.

Table 4.10: Mean and standard deviation of English and Mathematics marks

Attributes	Interval (X±1S)	High-value category	Average-value category	Low-value category
Mathematics mark	(0±2.45)	≥ 2.45	-2.45 … 2.45	≤ -2.45
English mark	(0±2.47)	≥ 2.47	-2.47 … 2.47	≤ -2.47

4.5 Preparation of Data for the Experiments

The category labels for the different values of the attributes are as depicted in Table 4.11. For instance the value "Positive" was given to refer to the high-value category for group work attitude.

Table 4.11: Category labels for each of the attributes

Personality Attribute	**Category labels**
Gender	Male, Female
Group work attitude	Positive, Indifferent, Negative
Interest for Mathematics	Interested, Indifferent, Uninterested
Achievement Motivation	High, Average, Low
Self Confidence	High, Average, Low
Shyness	Extrovert, Average, Introvert
Mathematics mark	Above satisfactory, Satisfactory, Below satisfactory
English mark	Above satisfactory, Satisfactory, Below satisfactory

The quantitative values of the attributes for each student were then changed into the above category labels. A sample of the resulting records for the same data records in Table 4.7 is shown in Table 4.12.

Table 4.12: Sample of the resulting records (number values replaced by category labels)

Gender	Group Work Attitude	Interest for Math	Ach. Motivation	Self Confidence	Shyness	English Performance	Maths Performance
Male	Positive	Indifferent	Medium	Medium	medium	Satisfactory	Satisfactory
Male	Indifferent	Indifferent	Medium	Medium	Introvert	Satisfactory	Satisfactory
Male	Indifferent	Interested	Medium	Low	medium	Satisfactory	Satisfactory
Male	Indifferent	Indifferent	High	High	Extrovert	Satisfactory	Satisfactory
Male	Indifferent	Interested	Low	Medium	medium	Satisfactory	Satisfactory
Male	Indifferent	Indifferent	Medium	Medium	medium	Satisfactory	Satisfactory
Male	Negative	Indifferent	Medium	Low	medium	Satisfactory	Above Satisfactory
Male	Positive	Indifferent	Medium	Medium	medium	Below Satisfactory	Satisfactory
Male	Indifferent	Indifferent	Medium	Medium	Extrovert	Satisfactory	Satisfactory
Male	Positive	Interested	Medium	Medium	Extrovert	Above Satisfactory	Satisfactory
Male	Indifferent	Interested	Medium	Medium	Extrovert	Satisfactory	Satisfactory
Male	Negative	Indifferent	Medium	Low	Introvert	Satisfactory	Satisfactory
Female	Positive	Interested	High	Medium	Extrovert	Satisfactory	Above Satisfactory
Male	Indifferent	Indifferent	Medium	High	medium	Satisfactory	Below Satisfactory
Female	Indifferent	Indifferent	Low	Low	Introvert	Below Satisfactory	Satisfactory
Male	Indifferent	Indifferent	Low	Low	medium	Above Satisfactory	Above Satisfactory
Male	Indifferent	Indifferent	Medium	Medium	Introvert	Below Satisfactory	Satisfactory
Female	Indifferent	Interested	Medium	Medium	medium	Satisfactory	Above Satisfactory
Male	Indifferent	Interested	Medium	Medium	medium	Above Satisfactory	Above Satisfactory
Male	Indifferent	Indifferent	Medium	Medium	medium	Below Satisfactory	Satisfactory
Male	Indifferent	Interested	Medium	Medium	medium	Satisfactory	Satisfactory
Female	Indifferent	Indifferent	Medium	Medium	medium	Satisfactory	Below Satisfactory
Male	Positive	Interested	Medium	High	medium	Below Satisfactory	Satisfactory
Male	Indifferent	Indifferent	Medium	Medium	medium	Satisfactory	Below Satisfactory
Male	Indifferent	Indifferent	Medium	High	medium	Above Satisfactory	Below Satisfactory
Female	Positive	Indifferent	High	Medium	medium	Below Satisfactory	Satisfactory
Male	Indifferent	Interested	Medium	High	medium	Below Satisfactory	Below Satisfactory
Female	Indifferent	Indifferent	Medium	Medium	medium	Satisfactory	Satisfactory
Female	Positive	Interested	High	Low	medium	Satisfactory	Satisfactory
Male	Indifferent	Interested	Medium	Medium	medium	Above Satisfactory	Satisfactory
Female	Indifferent	Interested	High	Medium	medium	Satisfactory	Satisfactory

At the end of measurement, it was observed that there were some inconsistent records where the values of the first seven attributes were the same for two or more students but different values were observed for mathematics performance. A java program was, therefore, written to count those inconsistencies in the data. An example of the inconsistencies observed is given below.

```
Male,Positive,Indifferent,Medium,Medium,Medium,Satisfactory,Below_Satisfactory
Male,Positive,Indifferent,Medium,Medium,Medium,Satisfactory,Satisfactory
Male,Positive,Indifferent,Medium,Medium,Medium,Satisfactory,Satisfactory
Male,Positive,Indifferent,Medium,Medium,Medium,Satisfactory,Above_Satisfactory
                                                                                  (2 inconsistencies)


Male,Indifferent,Indifferent,High,High,Extrovert,Satisfactory,Satisfactory
Male,Indifferent,Indifferent,High,High,Extrovert,Satisfactory,Below_Satisfactory
                                                                                  (1 inconsistency)


Male,Indifferent,Interested,Low,Medium,Medium,Below_Satisfactory,Below_Satisfactory
Male,Indifferent,Interested,Low,Medium,Medium,Below_Satisfactory,Satisfactory
                                                                                  (1 inconsistency)


Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Below_Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Above_Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Satisfactory

                                                                                  (2 inconsistencies)

Male,Negative,Indifferent,Medium,Low,Medium,Satisfactory,Above_Satisfactory
Male,Negative,Indifferent,Medium,Low,Medium,Satisfactory,Satisfactory
                                                                                  (1 inconsistency)


Male,Positive,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Below_Satisfactory
Male,Positive,Indifferent,Medium,Medium,Medium,Below_Satisfactory,Above_Satisfactory

                                                                                  (1 inconsistency)

Male,Indifferent,Indifferent,Medium,Medium,Extrovert,Below_Satisfactory,Satisfactory
Male,Indifferent,Indifferent,Medium,Medium,Extrovert,Below_Satisfactory,Below_Satisfactory
                                                                                  (1 inconsistency)
```

Figure 4.1: Sample showing inconsistent records in the data

A total of 54 such inconsistencies (10% of the total) were observed. This revealed that using those identified attributes and their corresponding values as experimental data would yield a maximum prediction accuracy of 90%, i.e., only 90% of the students would have their performance correctly predicted. We also observe that the minimum chance of accurate prediction of the level of performance of a student is 0.33. (33.33%)

This completes the presentation of the survey works undertaken for identification and measurement of attributes as well as the preparation of data for experiments reported in the next two chapters. While Chapter 5 deals with the experiments related to performance prediction, Chapter 6 presents the experiment on group composition.

CHAPTER FIVE

5. EXPERIMENTS RELATED TO PERFORMANCE PREDICTION

Experiments in relation to performance prediction were carried out in three phases. The first phase dealt with the development of the Bayesian network model to capture the relationships and dependencies between the attributes based on the results of the survey presented in Chapter 4. The second phase of the experiment focused on the evaluation of the network built in the first phase. Apart from collecting data on the various performance attributes, the work in this phase also involved the administration of exams prepared for the purpose of the assessment. The third phase of the experiment related to further improving the prediction accuracy of the model by exploiting domain expert knowledge to reinforce the learning capability of the model.

5.1 The Performance Prediction Model

5.1.1 Building the Bayesian Network

The belief network modelling software employed for the purpose of the experiment were the Bayesian network PowerConstructor and the Bayesian Network in Java (BNJ) software packages. Among the several publicly available learning algorithms, the Genetic Algorithm Wrapper for K2 (GAWK) in BNJ and the Three Phase Dependency Analysis (TPDA) in Bayesian network PowerConstructor were employed in developing the model.

Since results might be influenced by the selection of the test and training datasets, experiments were carried out by splitting the data into 3 partitions, i.e., a percentage split (3-fold) was used to partition the dataset into training and test data. Each partition, in turn, was used for testing while the remainder was used for training. This process was repeated three times for each of the learning algorithms and, at the end, every instance was used exactly once for testing. Finally, the average result of the 3-fold cross validation was considered.

The results from each of the learning algorithms are shown below.

(i) The Genetic Algorithm Wrapper for K2 (GAWK)

The training data sets provided to this software were prepared in a format that was acceptable for BNJ software. They were then fed to the software to yield the corresponding Bayesian network. Out of the 3-fold experiments, Figure 5.1 shows the best learned[24] Bayesian network using GAWK.

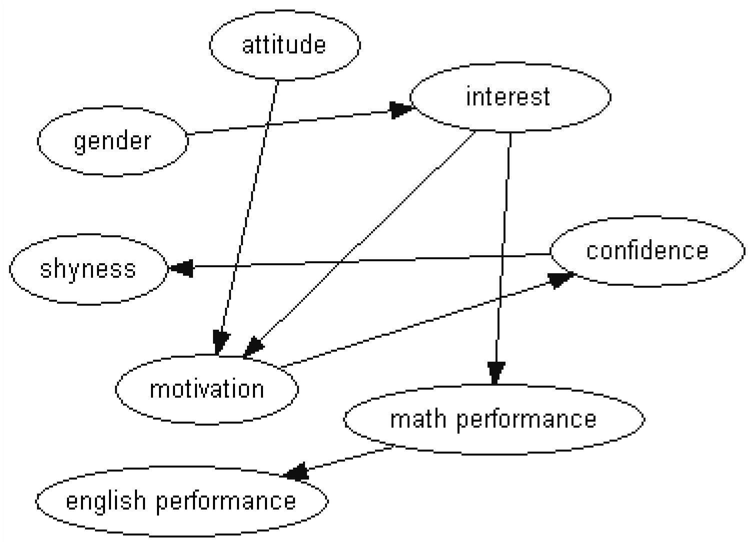


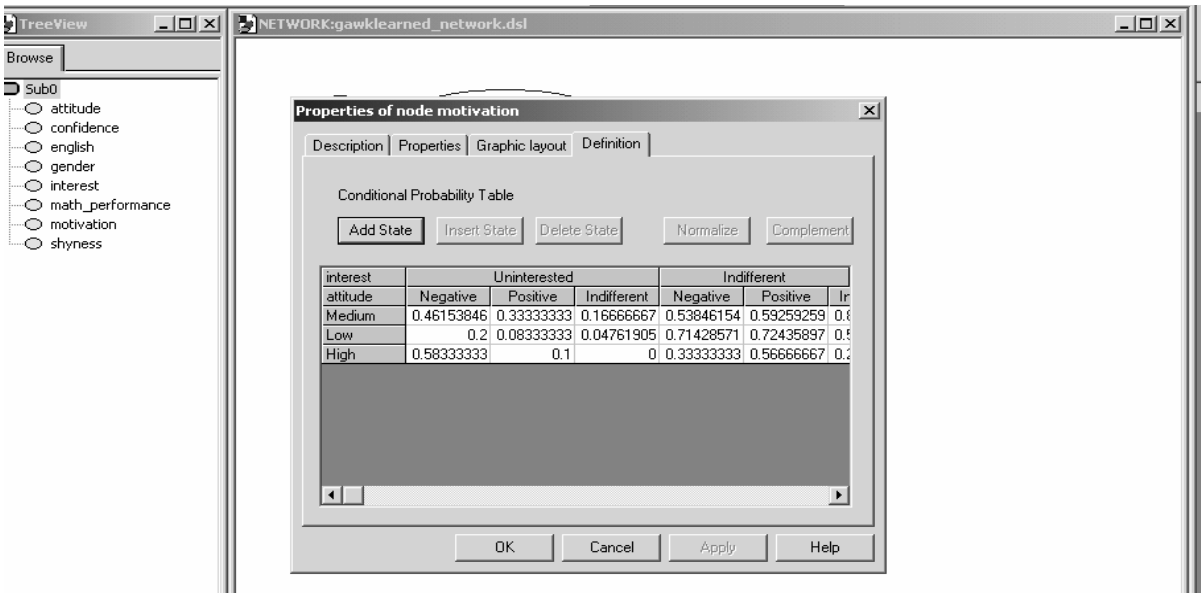
Figure 5.1: Performance Prediction Model (GAWK learned network)

Each node was described by a probability distribution conditional on its direct predecessors. Nodes with no predecessors are described by prior probability distributions. For example, node **attitude** (referring to group work attitude) was described by the prior probability distribution over its three outcomes (Positive, Indifferent and Negative). The other nodes were described by a probability distribution over their outcomes conditional on the outcomes of their predecessor. By observing values of attributes, captured from the model, one can compute the probability of performance .i.e.,

P(mathematics performance | gender, group work attitude, interest for math, achievement motivation, self confidence, shyness, English performance).

A sample of the conditional probability table for node "Motivation" from the network is shown in Figure 5.2.

[24] Best learned network is in terms of the prediction accuracy



interest	Uninterested			Indifferent		
attitude	Negative	Positive	Indifferent	Negative	Positive	Ir
Medium	0.46153846	0.33333333	0.16666667	0.53846154	0.59259259	0.8
Low	0.2	0.08333333	0.04761905	0.71428571	0.72435897	0.5
High	0.58333333	0.1	0	0.33333333	0.56666667	0.2

Figure 5.2: A sample of the conditional probability table (for node motivation)

(ii) The Three Phase Dependency analysis (TPDA)

For the TPDA algorithm, the training data sets were prepared as tables in Microsoft access database. The following figure illustrates the best learned network out of the 3-fold experiment.

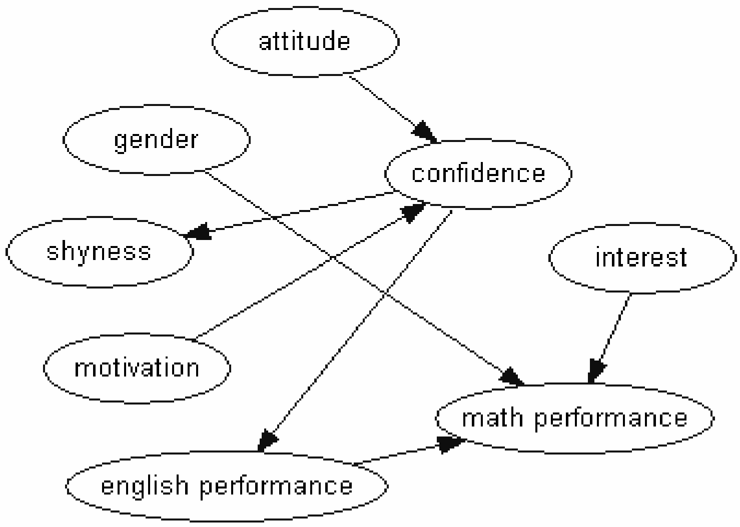


Figure 5.3: Performance Prediction Model (TPDA Learned network)

The TPDA learned network appears to be richer in explaining the dependencies between the various attributes than the GAWK learned network. For instance, in the TPDA learned network, Mathematics performance is observed to have three parents unlike the GAWK learned network where Mathematics performance has only one parent. A sample of the conditional probability table for node "Mathematics Performance" from the network is shown in Figure 5.4.

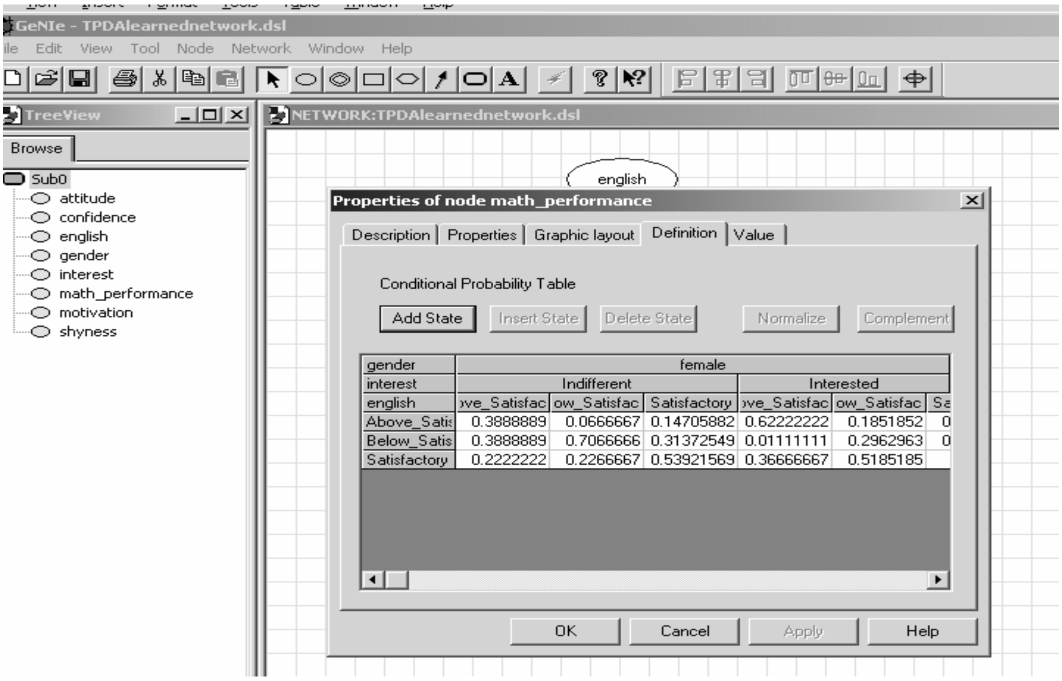


Figure 5.4: A sample of the conditional probability table (for node Mathematics performance

5.1.2 Testing the Bayesian Network

In order to test the performance of each of the learned networks, test data records were saved as text file. This file was then used to see the prediction performance of the model. A Java program was written to parse each of the test data records into the respective values of **gender, group work attitude, interest for mathematics, achievement motivation, self confidence, shyness and English performance.**

The Bayesian network in Java classes were used to load the learned network as well as carry out inferences. The Laurtizen-Spiegelhalter inference algorithm was employed for the purpose of predicting the prediction performance. Once the program loaded the corresponding learned network, the parsed information of each record was fed into the graph as evidence values. After consulting the model based on the evidence values, the program evaluated the probability values assigned for the three categories of mathematics performance. It then took the performance category having the maximum probability value and saved it together with the corresponding data record.

Upon completion of the prediction, it produced a confusion matrix by comparing the predicted performance level with that of the observed performance level. The confusion matrix (3-fold cross validation) from the two learning algorithms are shown in Tables 5.1 and 5.2.

Table 5.1: Output of the 3-fold cross validation using GAWK (confusion matrix)

	Predicted			
Observed	Below Satisfactory	Satisfactory	Above Satisfactory	Total No.
Below Satisfactory	14	23	4	41
Satisfactory	19	83	5	107
Above Satisfactory	4	7	13	24

The confusion matrix depicts that out of the total test records provided to the program, about 63.95% of the records were classified correctly. Moreover, as may be observed in the confusion matrix, the chance of classification of a "**below satisfactory**" performance category student into an "**above satisfactory**", which may be risky in the learning process was only 0.097 (9.7% of the below satisfactory students were predicted as above satisfactory).

Table 5.2: Output of the 3-fold cross validation using TPDA (confusion matrix)

	Predicted			
Observed	Below Satisfactory	Satisfactory	Above Satisfactory	Total No.
Below Satisfactory	12	26	3	41
Satisfactory	5	100	2	107
Above Satisfactory	10	4	10	24

As can be seen in Table 5.2, the TPDA learned network exhibited a 70.93% prediction accuracy. The chance of classifying a "**below satisfactory**" performance category student into an "**above satisfactory**", was only 0.073 (7.3% of the below satisfactory students were predicted as above satisfactory).

Because of its higher prediction accuracy, the learned network with the TPDA algorithm is used for further experiments.

5.1.3 Applying the Bayesian Network to Predict Performance

This section presents the program written to collect information from an individual student and predict the likely performance.

The program starts by presenting a form to enter **name, id, gender and English fluency** as depicted in the following screenshot (Figure 5.5).

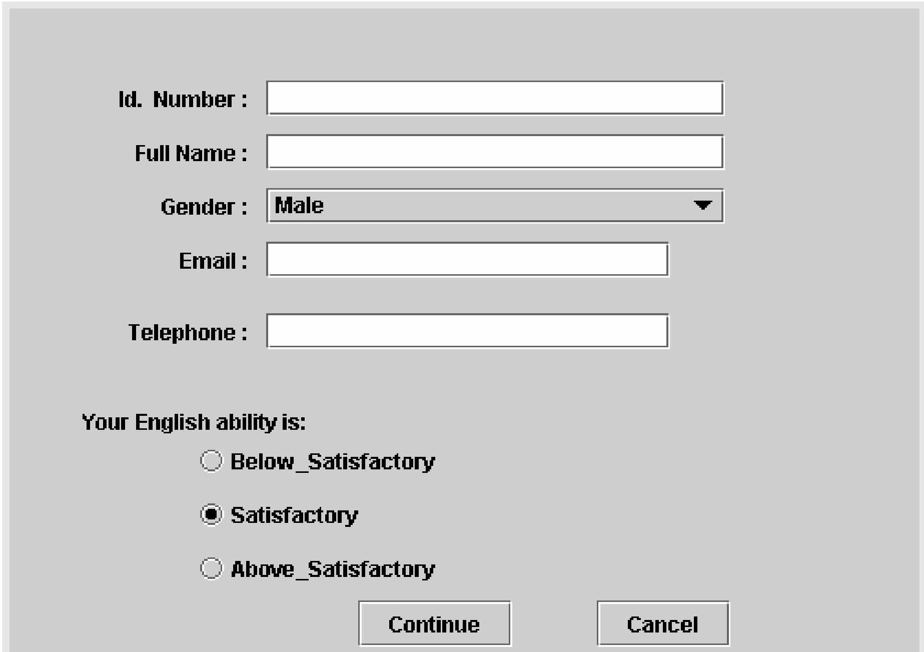


Figure 5.5: A screen to enter introductory information

After the student filled in the required information, another screen is displayed (Figure 5.6) where the student fills in his/her extent of agreement of the statements displayed for the purpose of measuring the various attributes.

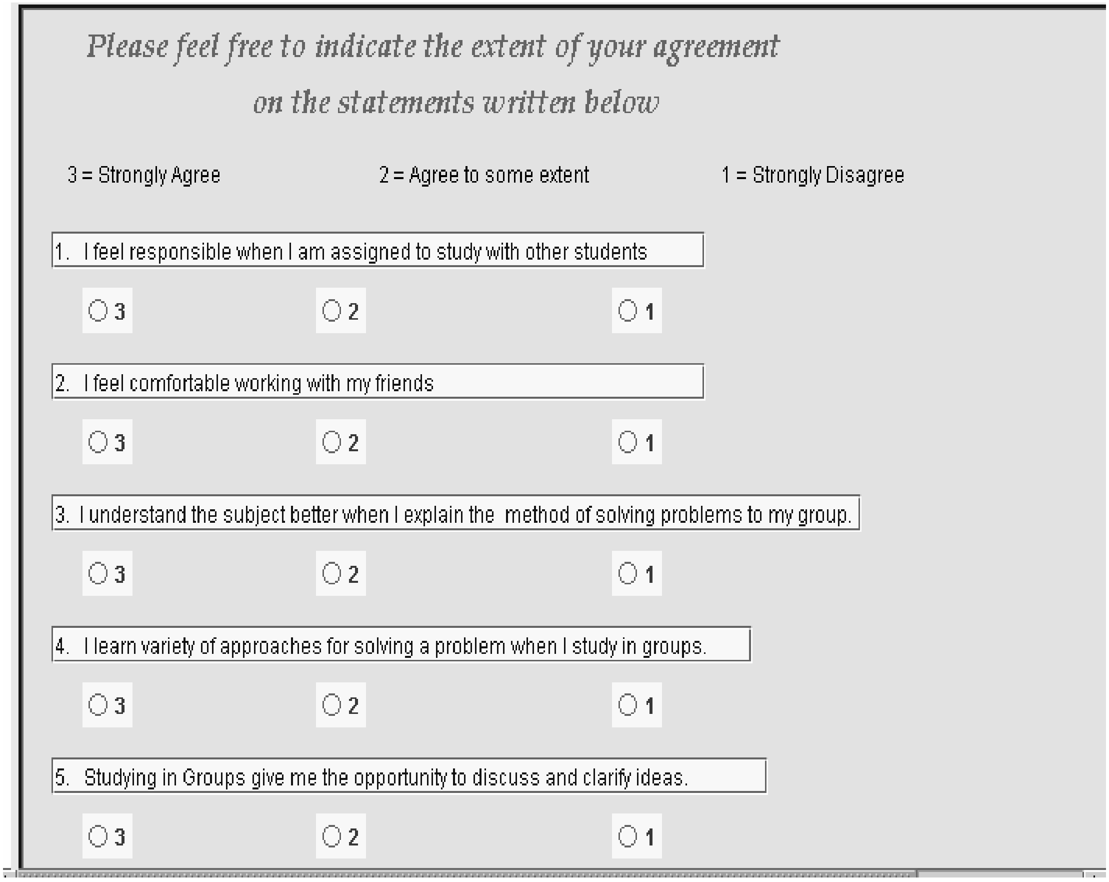


Figure 5.6: Sample of a screen shot showing attribute measuring items

After scoring the response of a student for each item, the scores of the items measuring a specific attribute are summed up. The program then makes a reference to rules described in Table 4.9 and Table 4.10 and assigns the corresponding category labels for each attribute (Table 4.11). Next, the program loads the learned network and feeds the values of the attributes as evidence values; it consults the inference algorithm for the probability of the student having “above satisfactory”, “below satisfactory” or “satisfactory” performance, it then takes the category with the highest probability and stores the information along with values of the other attributes.

5.2 Evaluating the Prediction Model in Real Environment

5.2.1 Student Samples

In order to further evaluate the prediction accuracy of the Bayesian network in a real classroom environment, the researcher particularly considered samples from 11^{th} grade students in the same school. The reason for considering 11^{th} grade students were: (i) at the time of conducting this evaluation, the students in grade 12 were busy preparing for the national exam; (ii) grade 12 students would not be able to participate in the group work experiment since they would be leaving the school by then, and (iii) the researcher also saw an advantage in testing the prediction model with a different set of students.

Information obtained from the school administration revealed that in grade 11, there were 615 students in the morning shift and 629 in the afternoon shift. While students from one section in the morning shift were made to involve in pilot testing of the exam questions, students from three sections in the afternoon shift were involved in testing the prediction accuracy of the model.

During the time of data collection, lessons were being delivered using a televised educational program from a central pool for all the classes with limited teacher interventions. This created a similar learning environment which minimized the influence from the variation that may have been introduced otherwise.

Details of the experiment carried out, are provided in the remaining sub sections of this chapter.

5.2.2 Inferences Made by the Prediction Model

Although the items to measure the five personality attributes were already automated, each individual student could not use computers, mainly because of shortage. Moreover, it was felt that misunderstanding of the statements in the English language, due to limitations in the English proficiency, might also affect the results. The approach used to address these two problems was to make students fill out the Amharic translated versions of the questions on paper, and then use assistants to feed the answers into the computer.

Students in three sections of the afternoon shift filled out the instruments. At the time of administering the instrument, there were a total of 55 students in Section 2, 50 students in Section 4, and 50 students in Section 6. In total, 155 students filled out the instruments.

The data protection and privacy issues were again explained to these students. The average time spent by the students while completing the instrument was 30 minutes.

Once the students filled in the data, it was fed into the program developed for the purpose of predicting the mathematics performance. As described earlier, the program made use of the prediction model built by TPDA learning algorithm and inference is made using the Lauritzen and Spiegelhalter inference algorithm. The performance categories of students based on the prediction model is shown below.

Table 5.3: Frequency distribution of predicted performance

Performance Level	Number of Students
Below Satisfactory	30
Satisfactory	97
Above Satisfactory	28
Total	155

5.2.3 Administration of Exam and Results

For the purpose of comparing the actual and predicted level of performance, the students who filled out the instrument, were made to write exams. The steps employed in preparing and administering the exam were as follows.

(i) Development of Questions

Topics, which have already been covered by the students, were selected from the already existing text book for grade 11. Care has also been taken so that the topics did not require a time consuming analysis – only those that could be worked out by simply applying rules were considered.

Once the topics were identified, questions were developed for each topic in consultation with mathematics instructors in the school. The questions were then given for comments to senior mathematics teachers especially those involved in preparing text materials for mathematics. Based on the comments, the questions were properly phrased and correctly set; ambiguity of notations, use of variables, phrases and brackets were avoided; each question was properly worked out and essential steps were identified.

Two sets of questions were then prepared, where each set contained 22 questions.

(ii) Questions tryout

A pilot test was carried out with the morning shift students of the Yekatit 12 Senior Secondary School. This was done in order to examine the difficulty level and the discrimination power of each question before the actual test was administered. The two sets of questions were administered on two different dates, since it was not possible to find suitable time where the students could answer all the questions in one day. For instance, using consecutive periods for the exams was not possible since the regular lesson in the next period which was delivered from a central distribution point would be interrupted.

The pilot test involved 115 students (students in two sections) who were not involved in the actual experiment. An average of 40 minutes was taken by the students to finish each set of questions on different dates. While correcting the question papers, it was observed that some students were not serious in answering the questions and left a good number of the questions unanswered. A total of 90 students attempted all questions and their answers were used to calculate the discrimination power and difficulty level of each question.

After ordering the marks of the 90 students in ascending order, 27% of the students[25] (i.e., 12 students with the highest and 12 with the lowest marks) were considered for further analysis. For these 24 students, the difficulty level and discrimination power[26] of each of the questions were computed. Questions with a difficulty level value of less than 0.54 and a discrimination power greater than 0.8 were considered for inclusion in the exam. A total of 15 questions were, therefore, taken out of the two sets of questions. Questions administered are attached as Appendix D.

(iii) Final Administration of Exam

For the purpose of comparing the actual and predicted level of performance, the same set of students whose performances were predicted, were made to write the final exam. The researcher with the help of the mathematics instructors at the school, administered the questions to these students.

While administering the exam, students were encouraged to attempt all questions. In addition to the general and specific directions given, oral instructions were given so that they clearly

[25] This percentage was taken in consultation with educational psychologists and based on available literature in Educational Psychology researches.

[26] Computations of the discrimination power and difficulty level were done by referring to materials on educational Psychology and discussions with experts in the field.

show the steps in answering each of the questions. The test was administered in a 40-minute period during regular class time. The average time spent by the students while answering the 15 questions was about 30 minutes.

At the time when the tests were administered, there were 47 students in Section 2, 48 students in Section 4 and 44 students in Section 6, making a total of 139 students. Some students who filled out the instrument were not available during the administration of the exam. These were 8 male and 8 female students. As observed from the instrument they had filled out, there was no particular feature missing which may bias the result of the comparison. Therefore, only those students who filled the instrument and took the exam were considered for the purpose of testing the prediction accuracy of the Bayesian model.

Each of the exam papers were then corrected out of 20 points. The mean mark was found to be 12 and the standard deviation 4 with maximum and minimum marks of 20 and 2.50 respectively. Based on the mean mark and standard deviation, those students who got greater than or equal to the sum of mean and standard deviation (≥ 16) were categorized as above satisfactory, those in between the difference of mean and standard deviation and the sum of mean and standard deviation (8…16) were categorized as satisfactory and those less than the difference of mean and standard deviation(≤ 8) were categorized as below satisfactory. The following table shows the number of students by performance category.

Table 5.4: Frequency distribution of actual performance – based on examination

Performance Level	Number of Students	Percentage
Below Satisfactory	24	17.26%
Satisfactory	89	64.03%
Above Satisfactory	26	18.71%
Total	139	100.00

5.2.4 Prediction Accuracy of the Model

The performance of each student based on actual exam results was compared with the predicted performance. The confusion matrix revealed a 66.18% accuracy, as shown in the following table.

Table 5.5: Accuracy of the performance prediction model (confusion matrix)

	Predicted		
Observed	Below Satisfactory	Satisfactory	Above Satisfactory
Below Satisfactory	16	7	1
Satisfactory	9	66	14
Above Satisfactory	0	16	10

As observed from the confusion matrix, there were a good number of students whose performance was predicted to be "**above satisfactory**", while their actual performance in the exam was "**Satisfactory**".

With the maximum prediction accuracy set at 90% this means that 23.82% of the students were wrongly classified. Further attempts made to improve the prediction accuracy are discussed in the following section.

5.3 Attempts to Further Improve the Prediction Accuracy

The first two attempts to improve the prediction accuracy emphasized on reducing the risk of misclassifying a student to a higher level of performance. The first attempt called for examining the computed probability values and use of a threshold value especially for those whose probability differences were not significant enough to classify a student to a higher level of performance. The second attempt was the introduction of a weighted probability value. The use of expert knowledge to modify the structure of the network was the third attempt. Each of these attempts is presented below.

(i) Use of Threshold Values

As stated earlier, one of the attempts made to improve the prediction accuracy was the introduction of a threshold value (D). If the absolute difference of probability values for two neighbouring categories is less than the threshold, then the lower category would be considered as the level of performance of the student. The rationale behind this is the assumption that the risk of classifying a "high achiever" as "low achiever" is safer than the vice versa. Accordingly, the following conditions were set:

Case (i)

- If the highest of the three probability values is the one with "**below satisfactory**", take the final prediction to be "**below satisfactory**".

Case (ii)

- If the highest of the three probability values is the one with "**satisfactory**",
  - Compare this probability value with the probability value corresponding to "**below satisfactory**".
    - If the absolute difference is less than D, take the final prediction to be "**below satisfactory**".

Case (iii)

- If the highest of the three probability values is "**above satisfactory**", then the system does the following in the order presented.
  - If the absolute difference between this probability and the probability of satisfactory is less than a certain threshold, take the final prediction to be "**satisfactory**".
  - If the absolute difference between probability of above satisfactory and probability of below satisfactory is less than a certain threshold, re-determine the final prediction to be "**below satisfactory**"

Although the attempt seemed reasonable, the difficulty lied in getting a justifiable threshold value. And hence, further attempts were not made along this line.

(ii) Use of Weighted mean

This approach was employed particularly to reduce the risk of misclassifying a "below satisfactory" student as "above satisfactory". With this approach, an attempt was made to replace the probability of "above satisfactory" by a weighted probability value(W_i). The weighted probability value(W_i) was computed as

$$\begin{aligned} P(W_i) = {} & \text{Relative frequency of low performers*P(Below satisfactory)} + \\ & \text{Relative frequency of satisfactory performers * P(Satisfactory)} + \\ & \text{Relative frequency of high performers* P(Above satisfactory).} \end{aligned}$$

The conditions set were:

- If the highest probability is that of "**below satisfactory**" then the final prediction is "**below satisfactory**".
- If the highest of the probabilities is that of "**satisfactory**", then the final prediction is "**satisfactory**"
- If the highest of the probabilities is "**above satisfactory**", replace the probability of above satisfactory with the weighted probability value and re-determine the probability of satisfactory and below satisfactory.
    - This was done with the assumption that if the probability of above satisfactory is much greater than "**below satisfactory**" or "**satisfactory**", it will remain higher even after the revision of the probabilities.

The exam results[27] presented earlier were used as weights and the weighted probability value was computed as:

$$P(W_i) = 0.17*P(\text{Below satisfactory}) + 0.64*P(\text{Satisfactory}) + 0.19*P(\text{Above satisfactory})$$

The probability values were then revised as follows.

Let P(A) denote the probability of above satisfactory;
P(B) denote the probability of below satisfactory;
P(S) denote the probability of satisfactory.

Then

- $P(A) = P(W_i)$ (new probability value for above satisfactory – replaced by weighted probability)
- $P(B) + P(S) = 1-P(A)$ (using basic axioms of probability)

Once the Probability of "above satisfactory" was replaced by the weighted probability value, probabilities of "satisfactory" and "below satisfactory" were re-determined based on proportional allocation.

$$\mathrm{P(B)} = \frac{\mathrm{P(B)}}{\mathrm{P(B)} + P(C)} * (1 - \mathrm{P(A)})$$

$$\mathrm{P(C)} = \frac{\mathrm{P(C)}}{\mathrm{P(B)} + P(C)} * (1 - \mathrm{P(A)})$$

[27] With the actual exam results, it was found that 17% have "below satisfactory" performance, 64% have "satisfactory" performance and 19% have an "above satisfactory" performance.

Once the probability values were re-determined, then the category corresponding to the maximum probability value was considered as the final prediction.

However, since the prior probability given for satisfactory performance is much higher (0.64), the experiment based on this concept made the predicted performance of almost all students "satisfactory" or "below satisfactory", hence reducing the prediction accuracy to 53.95%.

Since the prediction accuracy was reduced considerably, this concept was not found worth using.

(iii) Modifying the Structure of the Model Based on Expert Knowledge

A closer examination of the network may reveal some drawbacks, i.e., as it is learnt from sub-population of students, links may have been established between attributes which are independent in the general population; links may exist where direction should have been opposite from what appears in the network; links may exist where attributes are not directly related; an expert may also expect a variable to have several more parents than actually appearing on the network. This means automatic learning methods alone may not be sufficient. An option considered under such circumstances, was to reinforce the learning using the knowledge of the human/domain expert.

Therefore, the belief network (prediction model) was modified based on elicitation of the structure from experts and then obtaining the conditional probability tables from the existing experimental data. In order to do this, a survey, supplemented by discussions was conducted. The contents of the survey and the discussion guides were prepared based on the various structures observed from the already learned networks during the experiment. The experts were let to reason about the existence or not of: a direct cause or effect relationship between the attributes, indirect relationship, dependence or independence given certain conditions. These were also supplemented by available literature in the area of performance factors applicable to the local context.

A total of 56 experts from Mathematics Department in the same school, Departments of educational psychology, mathematics and English in Addis Ababa University, were involved in the discussion and surveys conducted. (See Appendix E and F for content of the survey, experts involved and responses).

After analyzing the responses of the experts, the network was modified accordingly as shown in Figure 5.7.

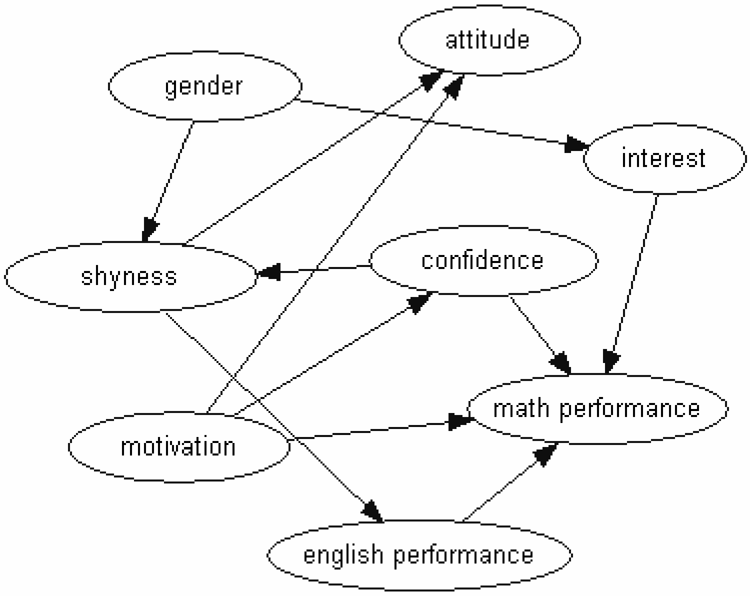


Figure 5.7: Network modified based on expert opinion

The major changes made in the modified network as compared to the original learned networks are as follows.

- In the original network we observed that mathematics performance had 3 parents, where as one more attribute (motivation) is included to have a direct influence over mathematics performance;

- In the original learnt network, we have gender directly affecting mathematics performance, where as in the modified network, gender affects mathematics performance only through interest for mathematics (the structure that was also observed in GAWK learned network);

- In the original learnt network, shyness has not been observed as a parent for any of the attributes while in the modified network, we find shyness affecting group work attitude and English performance;

- In the modified network, a new relation has been observed between achievement motivation and group work attitude.

Once the causal network was constructed with the help of human experts, it was then combined with quantitative estimates of conditional probabilities obtained from the database. The Bayesian network PowerConstructor system (TPDA-II algorithm) was employed to generate the conditional probability tables for each node in the network. One observes that with the modified network, the amount of information needed to specify each conditional probability table would be from 2 (where “gender” had no parents and two possible outcomes) to 3^4 (where “mathematics performance” had 4 parents and three possible outcomes). This meant, with 8 nodes the number of probability values computed was between 8 * 2 and 8 * 3^4.

This modified network was used to re-predict performance. The prediction accuracy was again compared with the actual performance as shown in the following confusion matrix.

Table 5.6: Confusion matrix from the modified network

	Predicted		
Observed	Below Satisfactory	Satisfactory	Above Satisfactory
Below Satisfactory	16	8	0
Satisfactory	3	77	9
Above Satisfactory	0	8	18

The confusion matrix revealed an accuracy of 79.85%. With the maximum prediction accuracy set at 90%, this means that only 10.15% (i.e., 90%-78.85%) of the students were wrongly classified.

This completes the discussion of the experiments carried out in relation to the performance prediction model. The next chapter is devoted to the discussion of the experiment conducted on the group formation process.

CHAPTER SIX

6. EXPERIMENTS RELATED TO FORMING HETROGENEOUS GROUPS

In this chapter the experiments carried out in relation to heterogeneous group composition are presented. The results of the works reported in the preceding two chapters (i.e., the results from the survey discussed in Chapter 4 and the output of the prediction model discussed in Chapter 5) formed the basis for the experiments presented in this chapter. In particular,

(i) The attributes identified to intervene with performance were considered in the group composition process;

(ii) The data prepared for the purpose of learning the structure of the network was used to test the algorithms developed to create the groups;

(iii) The exam administered to evaluate the performance prediction model was considered as pre-group work exam;

(iv) The same student samples who took the exam were involved in the group work at the second phase of the experiment, where grouping was made based on their predicted performance.

The first section of the chapter presents definitions and conceptual frameworks developed to mathematically formulate the heterogeneous group formation problem. Detailed accounts on the alternative algorithms developed, on the basis of the mathematical formulation, are presented in the second section of the chapter. The third section presents the experiments related to evaluation of the proposed grouping method in real classroom environment. The remaining sections are devoted to the discussion of the incremental version of the selected algorithm.

6.1 Conceptual Framework and Definitions

By applying the concepts of a vector space model, each student was represented in a multi-dimensional space by a vector whose features/components were made up of the values of personality and performance attributes. Definitions are given as follows.

(i) The Student Space Model

- Consider a student vector space, where each student is represented in a multi-dimensional space by a vector whose features are made up of the values of personality and performance attributes. $\mathbf{A_n(S_i)}$ is the value of the n^{th} attribute of student i. (In our case n =7)
- In other words, a student iss represented in the space by a point which corresponds to values for the 7 attributes, namely: group work attitude, interest for mathematics, achievement motivation, self confidence, shyness, English performance and mathematics performance. i.e.,

  $\mathbf{S_i}$(Attitude(S_i),interest(S_i), achm(S_i), selfconfidence(S_i), shyness(S_i), English(S_i) , math(S_i))

- Values of the attributes representing a student in space were weighted and mapped to numerical values. Since each of the seven attributes[28] had three possible values, the scores (numerical values) assigned for values of each attribute were: 1 for low category values, 2 for average category values and 3 for high category values.

  For instance, for

  S_1 (positive, indifferent, medium, low, extrovert, above satisfactory, satisfactory)

  the corresponding vector is represented by

  S_1(3,2,2,1,3,3,2)

(ii) Student-score

- Let a student-score for a particular student represent the total score computed as the sum of the scores for each of the attributes. In other words for a particular student j, the student-score is computed as

$$\textbf{Student-score} = \sum_{i=1}^{n} A_i(S_j) \qquad (6.1)$$

Where A_i (S_j) represents the score for a particular attribute A_i for a student j

[28] Since Gender is not given a weight(score), it was not considered in the score computation.

With the data collected for the purpose, the maximum value of the student-score is 21 when all attributes are in a high-value category and 7 when all attributes are in a low-value category. The figure below depicts the distribution of student-scores for the students considered in this experiment.

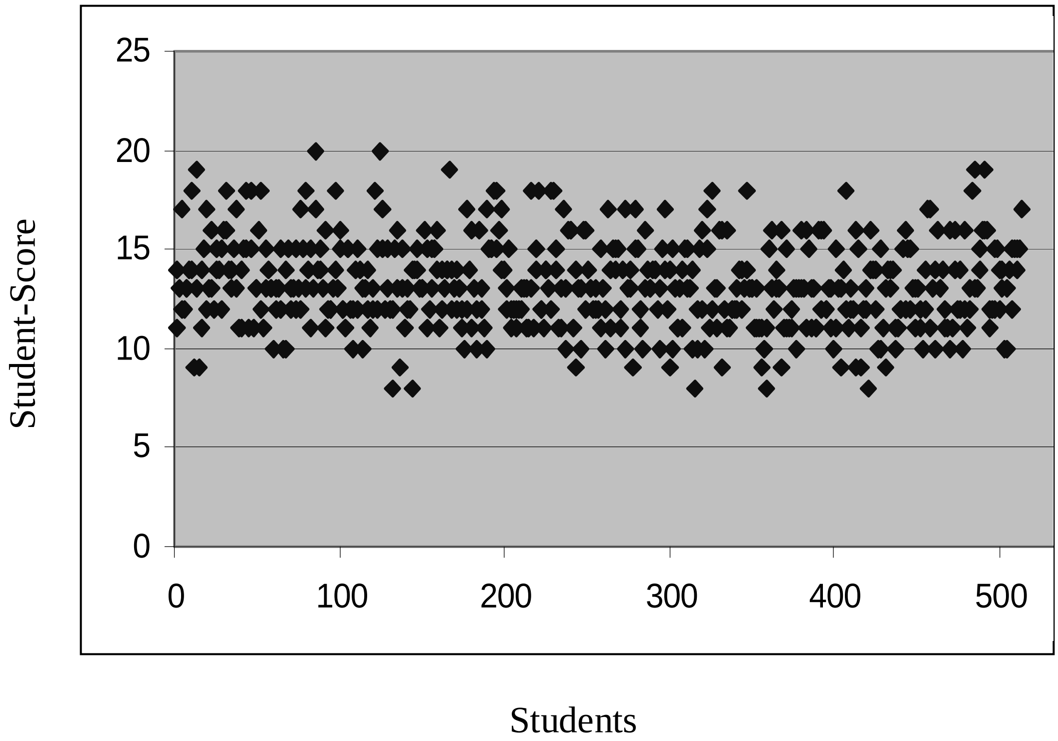


Figure 6.1: The student score distribution

(iii) Student-average-score

- Let Student-Average-score be a simple average computed from student-score of all students. i.e.,

$$\textbf{Student-Average-Score} = \frac{\left(\sum_{j=1}^{j=n}\sum_{i=1}^{i=t} A_i\,(S_j)\right)}{n} \qquad (6.2)$$

Where n is the number of students and t is the number of attributes.

(iv) The Difference-measure

- Let **Difference measure** Diff(S_1 , S_2) be defined as the distance between the vectors representing two students in space. Applying the Euclidean distance, this becomes

$$Diff(S_1, S_2) = \sqrt{\sum_{i=1}^{n} (A_i(S_1) - A_i(S_2))^2} \qquad (6.3)$$

- With n = 7, perfect similarity generates a value of 0 for students exhibiting no difference and a perfect difference yields a value of $\sqrt{28} \cong 5.29$. On the other hand, If one considers the student-score, perfect similarity generates a value of 0 for students exhibiting no difference and a perfect difference yields a value of 14 computed as 21-7.
- Such difference measures can be computed for all pairs of students (S_i , S_j) except when i = j. Note that for a total of t students, there are

  $p = \dfrac{t!}{2!(t-2)!}$ ways of making pairs.
- An **Average difference** can then be computed as

$$\textbf{Average-difference} = \frac{\sum_{i=1}^{i=t-1} \sum_{j=i+1}^{j=t} Diff(S_i, S_j)}{p} \qquad (6.4)$$

(v) Pair-threshold

- Let **pair-threshold** be defined as the lowest possible value of difference required to put two students in a specific group. The average difference between a pair of students is taken to be the pair-threshold.
  - In our case **pair-threshold = average-difference**. (refer to eqn. 6.4)

(vi) Heterogeneity

- Two students S_i and S_j are said to satisfy the heterogeneity requirement if

  $Diff(S_i, S_j) \geq$ pair-threshold

(vii) The Group-average

- Let Group-average = group-average(S_1 , S_2 ,…, Sg) represent the average score of a group computed from scores of the individual students included in the group. Where g stands for the number of students in a group, we have

$$\textbf{Group-average}(S_1, S_2, \ldots, S_g) = \frac{\sum_{j=1}^{j=g} \sum_{i=1}^{i=n} A_i(S_j)}{g} \qquad (6.5)$$

(viii) Group-threshold

- Let **group-threshold** be defined as the lowest possible value of the group average required to accept students for inclusion in the final group formation. In our case

  Group-threshold = Student-average-score (refer to eqn. 6.2)

- Group threshold is used for the purpose of selecting the students who would be included in the final group formation. i.e.,

  Make the group formation final **if Group-average > Group-threshold**

  where Group-threshold = Student-average-score

(ix) Group kernel (Group seed)

- **Group Kernel/Group Seed** is the student whom an incoming student is compared with based on the difference measure.

(x) Reasonably heterogeneous group

- A **reasonably heterogeneous** group refers to a group where student-scores in a group have a combination of low, average and high student-scores.

(xi) Criteria for effective group composition

- The main criteria considered while experimenting on the group composition were
  - initially all students are treated as outliers;
  - in a specific group, the difference measure between at least two students should be greater than the **pair-threshold**;
  - all students in a group should not have same low value for a given attribute (at least one should be either in high or average category for any one of the attributes);
  - the group-average should not be less than the **group-threshold;**
  - student-scores in a group should reveal a combination of low, average and high student-scores.

The first four criteria were considered in the process of creating groups, while the fifth criterion was used in the process of selecting the algorithm which yielded reasonably heterogeneous groups.

6.2 The Grouping System

6.2.1 Developing the Algorithms

In order to examine different approaches of heterogeneous group formation during the experiment, three algorithms were developed based on the conceptual framework outlined in the preceding section. Each of these are discussed below.

(i) Algorithm – 1: Considering the First Student as Group Kernel

This algorithm runs eight modules repeatedly before it reaches the final assignment of students to groups.

The peculiar property of this algorithm is that in the first module, the first student in the list is selected to serve as a group seed. Then all other students from the list are compared with the group seed to decide their assignment to the group represented by the seed (of course, by employing the difference measure). This process continues until the group size is filled. The whole process is then repeated by updating the list (excluding those students that have already been put into a group). The algorithm for beginning of group formation (first module) is attached as Appendix I(i).

The second module runs only if there are students who are not yet grouped (only if the outlier file still contains some students). The procedure is that first, it sequentially takes a student from the outlier file; it also sequentially selects a group which is not yet filled. It then applies the difference measure on the student from the outlier file and one member from the group. If the difference measure is greater than the pair-threshold with at least one of the students in the group, then the student from the outlier file is included in the group. The checking is done until each group is filled or there are no more students in the outlier file. The pseudo code of this algorithm is attached as appendix I(ii).

The third module by the name outlier-exchange runs only if there are still outliers. It performs a trial and error process starting from the 1st group. It temporarily removes a student from a group and replaces him/her by a student who is not yet grouped. It then checks whether the new student fits the join-requirement by applying the difference measure and comparing it with the pair-threshold. It actually tries until it finds a pair or until all students are checked.

If such a pair is found, the exchange is performed. This module does not change the size of the original outlier file (i.e., no reduction but simply exchange is done). See Appendix I(iii) for the pseudo code of the algorithm corresponding to this module.

The Fourth module creates new groups with the outlier file. It runs only if the outlier file contains students and it assumes that some students who were already grouped have now been exchanged to join the outlier students. The steps are actually the same us the algorithm in the first module.

The fifth module examines the final group candidates. It sequentially checks for all groups created. If all group members have low values for a specific attribute or if the group-average is less than the group-threshold, then the group is discarded and all members are put back to the outlier file. This module actually selects the final groups and drops those groups which do not meet the criteria specified. The pseudo codes of the algorithms corresponding to the fourth and fifth module are attached as Appendix I(iv).

The sixth module runs only if there are groups which are dropped when the final grouping is made. It simply repeats the first five modules. For the purpose of the experiment, these five modules were repeated until there were no more students who can be grouped together. The pseudo code of the corresponding algorithm is attached as Appendix I(v).

The seventh and eighth modules are concerned with outlier inclusion and finalizing the grouping process. In the seventh module, outliers are included on condition that a group size is not yet filled and if the group average becomes greater than the group-threshold after the inclusion of the student from outlier file. After checking all groups, if there are still outliers, the eighth module is run in order to append students from the outlier file sequentially to each group.

The pseudo codes of the algorithms corresponding to the seventh and eighth modules are attached as Appendix I(vi).

With algorithm-1, there were 119 groups made. A total of 476 students were grouped and 38 remained without any assignment. These were considered as the final outliers and were the ones considered in the seventh and eighth modules.

(ii) Algorithm – 2: Considering the Last Member as Group Kernel

This algorithm is a slight modification of Algorithm-1. The student that joined a group last (the latest or most recent in terms of joining the group) takes over the role of being a group kernel. In other words, the check for membership of a new/incoming student into a group is made by comparing the student with the one who joined the group last. This way, more than one student in a group will have a chance of being the group kernel (has an opportunity to pick a member and then transfer the opportunity to the one picked). The details of this algorithm are attached as Appendix I(vii). Other than this, the algorithm follows the same procedures as algorithm-1.

With algorithm-2, there were 122 groups made. A total of 488 students were grouped and 26 remained without any assignment to any group. As mentioned in the discussion of algorithm-1, these students were forced to join groups on a sequential basis based on the algorithm described in the seventh and eighth modules.

(iii) Algorithm – 3: Considering a Low Performer as Group Kernel

The third algorithm first searches for a student who has a below satisfactory performance in mathematics. This student remains the group seed until the group size is filled. In other words, the algorithm makes sure that there is always a student in a group who has a "below satisfactory" performance in mathematics. The details of this algorithm are attached as Appendix I(viii). The remaining steps followed are the same as Algorithm-1.

There were 112 groups made with algorithm-3, where only 448 students were grouped and 66 students remained without any assignment to any group. This algorithm actually made the smallest number of groups. Like the other algorithms, the students who were outliers were taken care of using the seventh and eighth modules.

6.2.2 Selecting the Best Algorithm

The concept of a reasonably heterogeneous group defined in Section 6.1(x) was employed in order to select the best algorithm. For instance, having a collection of all high or all low (or a combination of two high and two low) student-scores might not be the preferred composition as compared to having a group composition where student-scores have a combination of low, average and high student-scores (a reasonable heterogeneity).

The experiments carried out, in order to determine the algorithm which generated a higher number of reasonably heterogeneous groups, is presented below.

6.2.2.1 Group Average and Standard Deviation

Group Average

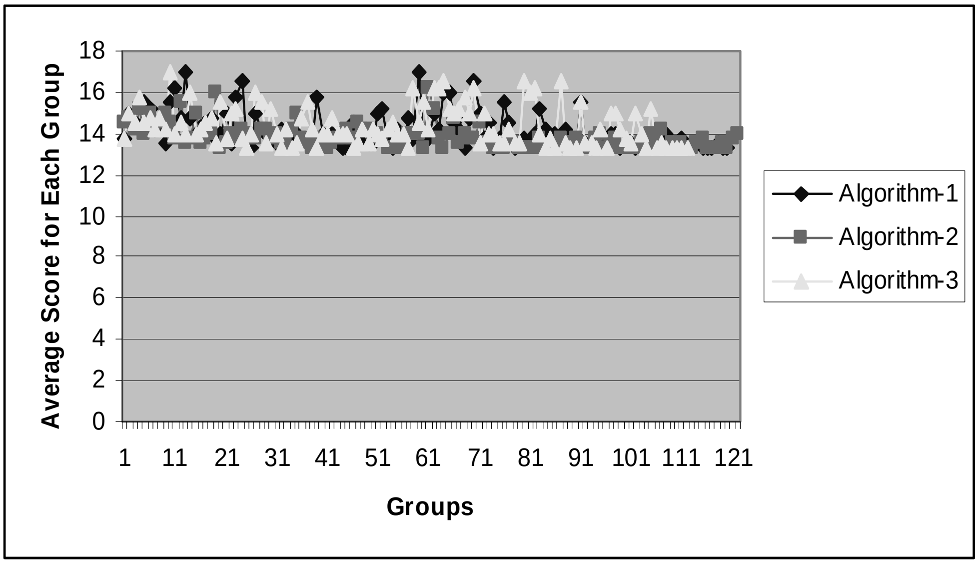
After computing the group averages for each of the groups created by the algorithms, the average of the group averages was computed. Table 6.1 is a summary of the results of the computation.

Table 6.1: Group average results of each of the algorithms

	No. of groups	Minimum group average	Maximum group average	Mean group average	Standard deviation of the averages	Coefficient of standard deviation[29] in the groups
Algorithm-1	119	13.25	17.00	14.11	0.87	6.17%
Algorithm-2	122	13.25	16.25	13.79	0.59	4.28%
Algorithm-3	112	13.25	17.00	14.35	0.97	6.76%

It is observed from the table that all the algorithms created groups with equal minimum group average.

[29] Standard deviation is only an absolute measure of variation. The Coefficient of standard deviation or simply the coefficient of variation (CV) is the relative measure of variation based on standard deviation and is given by $cv = \frac{\sigma}{\mu}$. When comparing between two observations, less CV indicates more consistency.



Graph 6.1: Group averages for each of the algorithms

As can be seen in Graph 6.1, the graphs generated from the group averages of the three algorithms were more or less identical.

The coefficient of standard deviation of the group averages computed for the three algorithms revealed that Algorithm-2 is the most consistent in terms of creating different groups with the same group average. However this was not a good enough justification to select this algorithm since there is no way of confirming whether a reasonably heterogeneous grouping has been achieved or not.

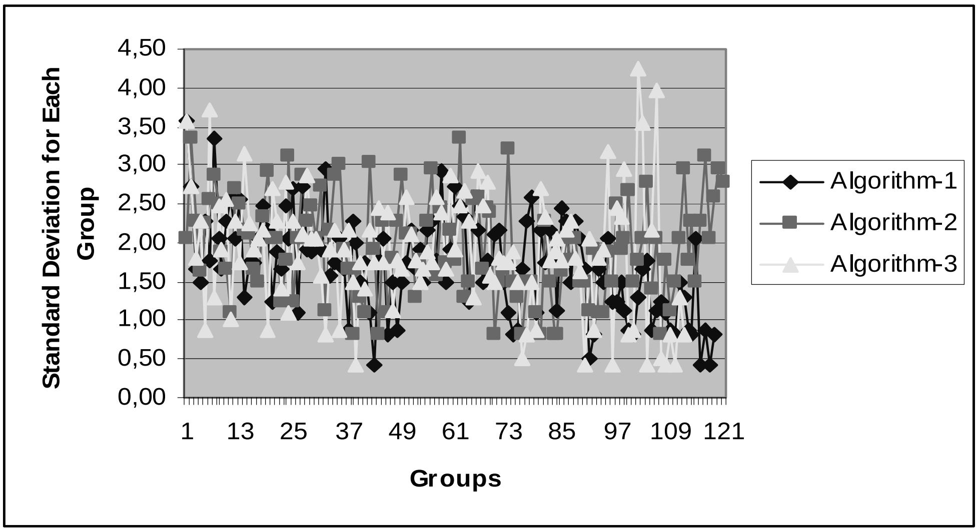
Measure of Variation

In order to measure how the students vary within each group generated by the three algorithms, the next step taken was measuring the standard deviation of the student-scores within the group. The following table is a summary of the standard deviation within the groups for all the algorithms.

Table 6.2: Standard Deviation of groups generated by each of the algorithms

	Number of groups	Minimum standard deviation obtained within groups	Maximum standard deviation obtained within groups	Mean std.
Algorithm-1	119	0.43	3.56	0.65
Algorithm-2	122	0.83	3.46	1.96
Algorithm - 3	112	0.43	4.24	1.89

The distribution of the standard deviation for each of the algorithms is shown in Graph 6.2. As can be seen from the graph, the standard deviation of each groups generated by the three algorithms are more or less identical.



Graph 6.2: standard deviation for each of the algorithms

Moreover, a high standard deviation might not reflect a reasonable heterogeneity within groups. For instance, if two students in a group have very high student-scores and the other two have very low student-scores, although the corresponding standard deviation will be high, it does not give us the expected heterogeneity within the groups. So in a sense, the standard deviation may be a misleading measure for heterogeneity.

After observing that neither the mean nor the standard deviation can be a good measure, there was a need to develop a mechanism to measure the Goodness of Heterogeneity(GH) or goodness of group composition. The next section describes the measure of goodness of heterogeneity developed for the purpose.

6.2.2.2 Goodness of Heterogeneity(GH)

(i) The Measure

As has been discussed above, the existing statistical measures (mean and standard deviation) were not good enough to select the best algorithm which created reasonably heterogeneous groups, the best being where a fair combination of low, high and average scores are observed. Hence a mathematical model to compare the goodness of heterogeneity was developed. The measure of goodness of heterogeneity was developed with the assumption that "**if there are**

high and low student-scores in groups, we also expect the rest of the students to be halfway between the maximum and minimum student-scores". This is defined as follows.

Let **Diffmax** be defined as the difference between the maximum and the minimum student-scores in the ith group.

Diffmax$_{(i)}$ = $\max scoreof\,(S_1,S_2,S_3,S_4) - \min scoreof\,(S_1,S_2,S_3,S_4)$ (6.6)

Let **Avediff** be the average score of the maximum and the minimum student-scores in the ith group

$$\textbf{Avediff}_{(i)} = \frac{\max scoreof\,(S_1,S_2,S_3,S_4) + \min scoreof\,(S_1,S_2,S_3,S_4)}{2} \tag{6.7}$$

The measure of Goodness of Heterogeneity is then defined as

$$\textbf{GH}_{(i)} = \frac{Diff\ \max_{(i)}}{1+\sum_j \left| Avediff_{(i)} - scoreof\ (S_{j(i)}) \right|} \tag{6.8}$$

Where $S_{j(i)}$ = the score of the jth student in group i, excluding the maximum and the minimum student-scores.

As explained above, the assumption that led to the above formula is that in a reasonably heterogeneous group, after taking the maximum and minimum student-scores, we expect the rest of the student-scores to lie half way between the maximum and minimum scores. This will make the absolute difference of the average difference(Avediff) and the rest of the student-scores the minimum possible. Where a reasonable heterogeneity is experienced, the numerator in equation 6.8 should be greater than the denominator hence yielding a high value[30] of GH. Similarly, a small value of GH indicates that the group is not reasonably heterogeneous.

It is trivial to show that GH is close to 0 when all students in a group have equal student-score; GH < 1 when there is unreasonable heterogeneity in the group (meaning student-scores are at

[30] In a reasonably heterogeneous group composition , we expect the summation part of the denominator closer to zero, and hence the whole part of the denominator closer to 1. (1 is included as a correction factor to prevent division by zero).

two extremes) and in reasonably heterogeneous groups, GH > 1. The greater GH the better the heterogeneity. To explain more about the measure of GH, some example cases are presented below.

Example case 1

Consider students in a group where two students have high scores and two students have low scores. - Eg. 20,20,8,8. The GH value in this case is 0.923.

Example case 2

Consider students in a group where three students have high scores and one student has low score. - Eg. 21,20,19,9. The GH value in this case is 1.2.

Example case 3

Consider students in a group where one student has high score and three students have low scores. - Eg. 20, 8,8,9. The GH value in this case is 1.

Example Case 4

Consider students in a group where student-scores are reasonably heterogeneous. - Eg. 9,13,17,21. The GH value in this case is 2.4.

(iii) Testing the Algorithms: Application of GH

The following are the GH values for each of the algorithms developed.

Table 6.3: Distribution of Goodness of Heterogeneity - groups created by Algorithm-1

GH Value	**No. of Groups**	**Percentage**
≤ 1	49	41.2%
1.2 – 2	57	47.9%
2.25 – 3	12	10.1%
> 3.00	1	0.8%
	119	100.00

Table 6.4: Distribution of Goodness of Heterogeneity - groups created by Algorithm-2

GH Value	**No. of Groups**	**Percentage**
≤ 1	38	31.1%
1.2 – 2	72	59.0%
2.25 – 3	8	6.6%
> 3.00	4	3.3%
	122	100.0

Table 6.5: Distribution of Goodness of Heterogeneity - groups created by Algorithm-3

GH Value	**No. of Groups**	**Percentage**
≤ 1	46	41.1%
1.2 – 2	48	42.9%
2.25 – 3	11	9.8%
> 3.00	7	6.2%
	112	100.0

As depicted in the tables above, a higher proportion of the groups with GH value less than 1 are observed in Algorithm-1 and Algorithm-3, as compared to algorithm-2.

A summary of the distribution of the GH values for each algorithm, is shown below.

Table 6.6: Summary of distribution of Goodness of Heterogeneity

	Algorithm-1	Algorithm-2	Algorithm-3
No. of Groups	119	122	112
GH ≤ 1	41.2%	31.1%	41.1%
GH > 1	58.8%	68.9%	58.9%

Algorithm-2 was then selected as the best algorithm as compared to Algorithm-1 and Algorithm-3 for the following reasons.

(i) the coefficient of standard deviation showed that the groups created by Algorithm-2 were the most consistent;

(ii) the number of groups with GH values greater than 1, are the highest in algorithm 2;

(iii) Algorithm-2 created the largest number of groups.

Accordingly, all further experiments on forming groups were carried out using Algorithm-2.

6.3 Evaluating the Grouping System in Real Environment

Another important experiment carried out as part of this research relates to the evaluation of the developed grouping system. This was carried out using the student samples considered in the evaluation of the prediction model. The experiment was conducted for two purposes namely, to evaluate the algorithm, and to test the existing claim that students placed in groups based on personality attributes are more likely to have a better performance than those placed randomly or on self-selection basis.

The setting of the experiment and related procedures are described below.

6.3.1 Setting of the Experiment

Group Formation: As in the previous experiment, the basis for the group work experiment were 11th grade students in Yekatit 12 Senior Secondary School in Addis Ababa. In order to avoid possible bias arising from school time and stream (area of specialization), the researcher selected those students in the science stream attending the morning shift. It was found that these students were conveniently grouped into three sections namely section 2, 4 and 6. Out of these students, the participants in the group work experiment were the ones who took the pre-group work exam. After explaining the purpose, the students were asked whether they would be willing to study in groups. Almost all showed willingness to study in groups with the exception of very few. The researcher took time to persuade those who were unwilling to participate in the experiment. As a result, it was managed to involve all students.

There were 47 students from Section 2, 48 students from Section 4 and 44 students from Section 6 making a total of 139 students.

A lottery method was employed in order to decide which grouping method to apply in each of the sections. Accordingly, section 2 were made to select their own groups (self-assigned groups. Students in section 4 were grouped based on their personality attributes using the group composition algorithm. Students of section 6 were randomly-assigned to groups, i.e., they were made to draw numbers (group labels: 1, 2, etc.) written on a slip of paper of the same size, colour and shape. Groups were then formed by putting members who picked the same labels together. With the recent introduction of the televised educational program from a central pool[31], all students who participated in the experiment attended the same type of lecture in the same format and by the same instructor. This has actually reduced the bias that may have been introduced otherwise.

The Head of counselling section and Mathematics teachers in the school cooperated in informing the students which groups they have been assigned to, when the group work begins and where to come for the group work.

Group work Environment: Before the group work actually started, orientations were given to students on how they should go about the group work, team-leadership and submission of the exercises. During the orientation, students agreed that having only one member in the

[31] This televised educational program from a central pool is transmitted to high schools all over Ethiopia.

group as a leader for the duration of the group work, was too much of a responsibility (particularly for students who had no prior experience in group work). Accordingly, the role of leadership started with one member in the group in alphabetical order, and each member of the group took turns on a weekly basis. The leader was responsible to report problems during group work, submit exercises and the weekly group report form.

After attending lectures in the morning, the students were made to meet regularly in the afternoons. A specific location was chosen for students during group work, where they work in their groups of four or five for about six hours per week (pictures of students during group work is attached as Appendix J). For the purpose of the experiment, worksheets were prepared based on the text and lectures in class. Although the students were free to discuss the problem in any language[32] they feel comfortable with, the worksheets were prepared in the English language since the lectures were also delivered in the same language. The group work generally consisted a weekly cycle of activities as follows:

- A worksheet consisting of exercises was distributed at the beginning of each week;
- Each team discussed the worksheets distributed prior to working on the exercises;
- Each team then worked together on problems, compared answers, and helped each other with difficult problems;
- In cases of difficulty, each team consulted mathematics instructors who regularly visited the group work;
- At the end of the week, the group leader submitted the answers;
- Submitted answers are corrected and returned to the team.

Supervisions: Since all the three sections met mostly at the same time, three monitors recommended by the head of the counselling section were hired by the researcher in order to supervise the group work. These monitors were assigned to each section based on their time convenience. Their tasks were mainly to assist the researcher in taking attendance, attending to problems of students during group work, controlling disciplinary problems and collecting weekly reports submitted by the group leaders.

In order to make sure that students actually participated in the group work, they were encouraged by their Mathematics instructors to help each other and actively explain to one another how to solve the problems in the worksheet. They were also told that in order to be

[32] In almost all the cases, the students were observed to discuss the problems in the Amharic language with little use of the English language.

rewarded, they have to ensure a high group score for the exam after the group work. This was partly the method suggested by Slaving (1983b) in order to motivate working in groups.

Duration: The group work took place from February 16 - April 16, 2004. A total of 48 hours were spent in group work before the post-group work exam was administered.

Problems Observed: One of the problems experienced was that students simply lacked the experience on how to function and behave in a group setting. During the first two weeks, there was a problem of lateness and absenteeism in some groups. Students had difficulty taking turns in leadership. There were however, groups that worked well from the very beginning and seemed to be able to get along well.

It was also observed during the group work that some students came unprepared. They just sat in the sessions and did not try to actively participate in the group. There were also some interruptions as a result of exams scheduled by other instructors during group sessions.

Overall, there was a noticeable progress in the activities of students during the group sessions. The weekly report form collected from students did not indicate any discomfort resulting from working together.

6.3.2 Administering Post-Group Work Exam and Questionnaires

Post-Group Work Exam

In order to evaluate the change in performance, post-group work exams were administered after completion of the group work. The questions were more or less similar in nature to what the students have been working on, during their group study. Before the administration, the test questions were discussed with the mathematics instructors at the school for their appropriateness to measure performance after group work (Exams administered are attached as Appendix G).

As indicated earlier, there were a total of 139 students for the pre-group work exam. However, at the time of administering the post-group work exam, 10 students who participated both in the pre-group work exam and group study, were not available. Thus only 129 students were considered for further analysis. The maximum mark out of possible 20, was found to be 20 and minimum 7.50. The mean mark was 14.5 and the standard deviation 3.76.

For the purpose of comparison, the mean and standard deviation of the results of the pre-group work exam were taken to categorize the marks into the three levels of performance. Accordingly, those students who obtained marks >=16 were categorized as "above satisfactory", those in between 8–16 were categorized as "satisfactory" and those with marks <=8 were categorized as "below satisfactory". The following table shows the total numbers found in each category.

Table 6.7: Performance after group work

Performance Level	Number of Students	Percentage
Above Satisfactory	52	40.3%
Satisfactory	74	57.4%
Below Satisfactory	3	2.3%
Total	129	100.0

Comparison of the pre- and post-group work exam results (Tables 5.4 and 6.7) showed that after the group work, there was an increase in the number of students who were in the category of "satisfactory" and "above satisfactory". There was also a considerable decrease in the number of students who were in the "below satisfactory" category.

Post- Group Work Questionnaire

Students were also asked to complete a group evaluation survey at the end of the group work. Survey contents mainly included opinion of students on group formation, how well they worked together and improvement in performance. In order to control the misunderstandings that may arise from language barriers, the survey contents were prepared in Amharic language. The English version is attached as Appendix H.

The data collected from the survey was then organized and analyzed using the SPSS package.

Further discussions on the results of the experiments, the statistical tests applied, and the feed back received from students are presented in Chapter 7.

6.4 Adding More Features to the Selected Algorithm – Incremental Version

In order to be able to group students on first come, first served basis, or depending on the availability of student records, an incremental version of the selected algorithm (where each incoming student became the group kernel) was also developed. This algorithm did not require the whole data set to be available in advance. As the pair threshold and group threshold could not be obtained without the whole dataset, the summary values obtained from the previous experiment were used.[33]

The algorithm performs the following tasks before assigning a student to a group.

(i) Collect required information from a student and predicts mathematics performance;

(ii) If the file is empty, the student will be waiting for other students to join;

(iii) If the file is not empty, the system compares the incoming student with each existing group where the size is not yet filled. The comparison is made with the student who joined the group last;

(iv) The student is placed in a group where the computed difference measure is greater than the pair-threshold and where the greatest difference measure is observed;

(v) The latest incoming student will serve as the group kernel;

(vi) Before assigning the fourth student to a group, the system checks the following

   i. the difference measure is greater than the pair threshold(**d**);

   ii. the value of the group average as the student joins the group, i.e., Group-Average should be greater than the Group-Threshold; and

   iii. an attribute should not have the same low value for all members;

VII. If all the three conditions are fulfilled then the student will join the group. Otherwise, the program looks for another group where a student could join or a new group will be created where the student becomes the group kernel waiting for other students to join.

The pseudo code of this algorithm is attached as Appendix I(ix).

[33] The use of previous threshold values is based on the assumption that if there is no complete dataset available to calculate the pair threshold and group threshold, other thresholds already used for other group composition of similar cases may be used as substitutes.

Checking Group Members: The student first enters his unique identification. The software then searches the group label having the same id. It then displays the name, gender, email and telephone numbers of students who are assigned in the same group as the student.

Testing: The 4th year students of the Department of information science were engaged in testing the proper running of the algorithm. Based on the information entered from 25 students, the program is found to run correctly.

6.5 Stability and Robustness of the Incremental Version

(i) Stability

The following steps were followed in relation to a statistical test applied to prove the stability of the incremental group composition algorithm as compared to the batch processing algorithm.

(i) The list of students was arranged in alphabetical order

(ii) The incremental group composition algorithm was applied to create groups; In this case, each record was fed to the program on a sequential basis. There were a total of 12 groups formed.

(iii) The goodness of heterogeneity for each group generated by the incremental version was calculated;

(iv) The mean goodness of heterogeneity as well as standard deviation were computed (See the 1st row, column 3 and 4 of Table 6.8);

(v) The same list of students in (i) were grouped using the batch processing algorithm

(vi) The goodness of heterogeneity for each group generated by the batch processing algorithm was calculated;

(vii) The mean goodness of heterogeneity as well as standard deviation were computed (See the second row, column 3 and 4 of Table 6.8);

(viii) Since the population variance was not known, its estimate (pooled variance) was computed as

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$ (See column 5 of Table 6.8)

(ix) The null hypothesis that there is no difference between the mean of Goodness of heterogeneity of groups formed by the incremental group composition algorithm and the batch processing algorithm (i.e., H_0: $\mu_1 = \mu_2$) was tested with a significance level of 0.05.

(x) The test statistic (t value) was given by

$$t = \frac{mean_1 - mean_2}{\sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}}$$

The summary table is shown in Table 6.8.

Table 6.8: Summary of data used to test difference between the two means.

Algorithm	No. of Groups formed	Mean goodness of heterogeneity	S.D	Pooled Variance	t value	t critical at α=0.05 (95%)
Incremental version	12	1.52	0.49	0.66	1.040	2.07
Batch Processing	12	1.80	0.80			

Since the calculated t-value is less than the critical value, we can safely claim that there is no difference between the two means, implying that both algorithms generate similar groups with almost same goodness of heterogeneity. This also confirmed that the incremental version is stable enough to be applied in cases where data records for all students are not available in advance.

(ii) Robustness

One method used to see the robustness or consistency of the incremental group composition algorithm was using the algorithm on the same data records but in different orders. The steps applied were as follows.

(i) Data ordering – the same data set was ordered with five different arrangements: first, alphabetically in ascending order; second, with a lottery method; third, with a systematic sampling where every 5th was considered for ordering; fourth, every 3rd was considered for ordering; and fifth, alphabetically in descending order;

(ii) The Incremental group composition algorithm was applied for each data set;

(iii) GH values for each group generated in each data set was calculated and the mean goodness of heterogeneity was computed as shown in the following table;

Table 6.9: Mean of the GH values generated for groups with different data orders

	Mean of the GH values				
Group	Alphabetical (Ascending)	Simple random (Lottery)	Systematic (k=5)	Systematic (k=3)	Alphabetical (Descending)
Mean	1.52	1.58	1.59	1.59	1.39
No. of groups formed	12	12	12	12	12

(iv) The hypothesis that the mean of goodness of heterogeneity for all groups is the same was then tested against the alternative hypothesis that at least one of the means is different;

(v) The analysis of variance (ANOVA) between the mean goodness of heterogeneity generated by different order of data is shown below.

Table 6.10: Summary table of one-way ANOVA

Source	Degrees of Freedom(DF)	Sum of Squares[34]	Mean Squares[35]	F[36]
Variation Between groups	4	0.346	0.0865	0.179
Variation Within groups	55	26.62	0.484	

The result of the analysis revealed that there is no significant difference ($F_{(4/55)}$, 0.05 = 2.53) between the mean goodness of heterogeneity of the five differently orders student data sets further confirming that the algorithm is robust.

---

[34] The column labeled **Sum of Squares** describes the variability in the goodness of heterogeneity values.

[35] The column labeled **Mean Squares** are the Sums of Squares divided by the corresponding degrees of freedom.

[36] The column labeled F is the ratio computed as $\frac{\text{MS(Variation between groups)}}{\text{MS(Variation within groups)}}$

- The idea is if the variation between groups is much higher than the variation within groups, it may be reasonable to reject the hypothesis. The ratio follows an F distribution with 4 (5-1) and 55 (60-5) degrees of freedom.
- The number 60 refers to the total no. of groups formed using all types of data ordering.
- The number 5 refers to the number of data ordering methods used.

CHAPTER SEVEN

7. RESULTS AND DISCUSSION

In this chapter, we will present the results and discussions of the experiments conducted in a manner that will address the research questions established at the outset. To render a meaningful flow, the order of presentation followed in the preceding chapters has been maintained. Accordingly, the results and discussions on the identification and measurement of important attributes are presented first. In the second section, discussions are made on the performance prediction model. The third and fourth sections present discussions related to group composition and feed back from students respectively. The chapter concludes by pointing out some of the implications of the results of this study in the field of education.

7.1. Identification and Measurement of Attributes

As detailed in Chapter 4, the environments considered for the experimentation were the Mathematics subject and Ethiopian high school students in the preparatory (for tertiary level) program.

Based on a survey conducted for the purpose, the attributes “gender”, “group work attitude”, “interest for mathematics”, “achievement motivation”, “self confidence”, “shyness”, “level of English performance” and “level of mathematics performance” were identified as group composition factors. In this connection, it is relevant to note that according to the survey results, the first seven attributes were considered as factors that intervene with the level of mathematics performance. The results of the survey were also in accord with what was suggested in the literature in relation to group composition and performance factors.

As has been observed, the collected data from the outset revealed some inconsistencies where all the seven attributes were the same for two or more students but a different value for mathematics performance. This has made the maximum prediction accuracy, that can be obtained from the performance prediction model, only 90%.

A number of reasons could be mentioned as causes for some of the inconsistencies observed in the data. One reason might be that students were not prepared enough to appreciate the use of such research works and were not serious and honest in filling out the instrument. Another reason may be related to cultural issues. For instance, in the Ethiopian culture, it is not a commonly desired behaviour to talk of oneself highly and to make the feelings of one known publicly. In fact, in their upbringing (the parenting style mostly being authoritarian), children are required to keep low profile, be reserved and particularly not to be boosters of their good deeds. The norm is rather to express a lower opinion than is probably deserved of one's own ability, knowledge, skill, success, etc. Often, others are expected to talk on one's behalf. In certain conditions, there is also a fear of falling victim to one's full disclosure or assertiveness. Lack of confidence might have also resulted in not providing accurate data about themselves. What all this means is that, the possibility that cultural and social influences which inhibit students from giving answers on their true self, even where they have performed very good in their Mathematics exams, can not be overruled.

Taken together, some of the above mentioned factors might have caused the inconsistencies observed in the data and hence affected the prediction accuracy. Addressing these problems, in shorter term, may require instructors and school administrators (as well as investigators in the area) to put more efforts in preparing students for such work/survey through conducting appropriate orientation and persuasion/encouragement. For the long term, educationists need to work cautiously and systematically in addressing the issues related to cultural and social influences on self assertiveness and confidence building.

Given more time and resources, further exploration and inclusion of additional attributes might also contribute to better prediction accuracy. For instance, this study has not considered family economic and educational background and parenting style which were discussed as additional factors that intervene with performance. However, getting such information as economic background might be difficult under the existing circumstances in Ethiopia. For instance, information systems that might help in this regard (such as credit and banking systems, tax system, social security information, etc.) are either absent or premature. Cultural aspects of the sort discussed above also inhibit parents to openly/publicly declare their income. To this end, special mechanisms need to be devised to collect information for use in such research works. Particular to mathematics, other factors such as absenteeism from school and math anxiety might also need further consideration to modify the network and improve the prediction accuracy.

Further more, given stable economic and political situation, religion and ethnic background might be considered in group composition. This might allow for more flexibility of the group work. As such, students with the same religion could have the same worshipping time without disrupting the group work.

Since the experiment was conducted in a high school in Addis Ababa, it was convenient to assume that all students in Addis Ababa speak fluent Amharic, although they have their roots in different racial, national or tribal groups. If and in the event where such research work takes place in college/university environments where students come together from all over Ethiopia, one may need to consider ethnic background of students including native language in the group composition.

7.2 Performance Prediction Model

As reported, a system has been developed to predict performance using some personality factors that are said to affect performance. The approach was more of studying the dependency relationships among the attributes identified based on collected data and that of finding a network that best matches the data.

As it has been indicated during the experiment, the original performance prediction model where both the structure and conditional probability tables were learnt from the data records of students, has a performance prediction accuracy of 66.9%. The beliefs of the network about the situation have also been modified based on the inputs from domain experts. In this connection, one should note that a major advantage of Bayesian Network is that the knowledge/belief in the network is represented in a way that agrees with human intuition so that the network structures can be easily understood and that additional domain knowledge can be easily incorporated.

Based on the modified performance prediction model, it is observed that 79.8% of the students involved have their mathematics performance correctly predicted. Because of the inconsistencies observed from the measurement process, it was indicated in Chapter 5 that the maximum prediction accuracy was 90%. We may, therefore, claim that only 10% of the students were wrongly predicted.

Network structure learned from data

Although the developed network model seems to be confirming the claims made by different researchers in relation to factors affecting performance, some of the results in both prediction models were quite interesting as they were not fully in accord with what one may have expected on the basis of the discussion in section 2.1 of Chapter 2. For instance, it has long been argued that gender affects both language ability and mathematics performance. However, in this case, no relation was shown to exist between gender and English performance. On the other hand, a direct relation between gender and mathematics performance was observed. In fact, the only attributes that directly affected mathematics performance were gender, English performance and interest for mathematics. As could be seen from the network, mathematics was independent of its non-descendant self confidence, attitude and motivation given English performance. One may also note here that absence of arcs indicated independent events.

Network structure modified based on expert opinion

Indeed, there was a slight difference between the node ordering based on expert opinion and the automatic ordering produced from the learned data. The node ordering based on expert opinion showed that performance in mathematics is independent of its non-descendant gender given interest for mathematics. A new causal effect where shyness directly affected English performance and group work attitude was also added. One other attribute, “achievement motivation”, which directly affected Mathematics performance, was added to the network. Performance in mathematics was independent of its non-descendant “extent of shyness” given English performance. Group work attitude was independent of gender and self confidence given shyness. It should be noted here that more observations of such sort could be made from the network. We observed that as compared to the network which was automatically learned from the data set, the modified network is richer in explaining dependencies and relationships which might also have contributed to better prediction accuracy.

In order to generalize this observation for the entire student population and to be of further use to researchers in the area, it may be necessary to interpret these results in terms of probability of occurrence of events. From the generated conditional probability tables[37], probabilistic assessment of the dependencies in the network reveal that there is a chance of 0.54 for male students to be interested in mathematics, while for female students the chance is only 0.35. Female students with an “above satisfactory” English performance and who are interested in mathematics have a higher chance of being in above satisfactory level of performance in mathematics as compared to males with the same English performance and interest in mathematics - (62% for females and 51% for males). On the other hand, even with “above satisfactory” performance in English, the chance of being at above satisfactory level in Mathematics for female students with no interest for mathematics is lower (0.18) as compared to males with the same values of the attributes (0.23).

Further more, any student with high self confidence, above satisfactory English performance and high motivation and who is interested in mathematics, would have a 70.5% chance of being “above satisfactory” in mathematics and only a 2% chance of being in a “ below satisfactory” category. A student who is in low-category for all the four attributes has only

[37] Please note that it is not possible to discuss all probability values generated from the network.

0.7% chance of having above satisfactory performance in mathematics and a 78% chance of being in low-category performance.

In summary, by using the prediction model developed, we have managed to avoid the requirements of written exams to determine level of performance in mathematics for the purpose of forming groups. In stead of obtaining performance data from exams administered at the beginning of group formation, the solution developed can be of use to predict level of mathematics performance from values of other personality attributes.

7.3. Group Composition

As detailed in Chapter 6, for the purpose of creating heterogeneous groups, the student population was mapped to a mathematical model. Three algorithms were then considered for the purpose of creating the groups based on the mathematical model. A measure of goodness of heterogeneity was also developed in order to select the best algorithm. Based on the selected algorithm, the incremental version was also developed. The first approach (based on a batch-processing algorithm) is used when there is a need to create grouping for a set of students (for instance, students in a class). The other approach (incremental version) does not need the whole data set in advance.

In order to test the effectiveness of the algorithm developed, three grouping methods were employed to allocate students into groups, namely: random assignment, self-assignment and program-assignment. After working in groups for about 8 weeks (48 hours), a post-group work exam was administered. Results are discussed in Section 7.3.1 below.

7.3.1 Comparison of Pre- and Post- Group Work Exam Results

(i) Change in Performance

The following is a summary table comparing the two exam results (pre- and post-group work).

Table 7.1: Summary of exam results of the pre- and post- group work

	Maximum	Minimum	Mean	Standard dev.	Coefficient of variation
Pre-group work exam	20	2.50	12	4	33.3%
Post group work exam	20	7	14.5	3.76	25.9%

According to the summary results, the mean mark of the exam results for the post-group work is higher than the mean mark of exam results for the pre-group work. Moreover, with a coefficient of variation of 25.9%, exam results of the post- group work showed more consistency as compared to the exam results of the pre-group work.

The test for significance of correlations made at $\alpha = 0.05$ ($r = 0.507$, $p < 0.05$), revealed that there is a highly significant correlation between exam results of pre-group work and post-group work, i.e., students who did well on the pre-group work exam also did well on the post-group work exam.

The paired samples T-test was also applied to test whether there is a significant difference between the two exam results. The following is a summary table.

Table 7.2: Paired Samples Test

	Mean	Standard deviation	Std. Error of mean	Value of z statistic	P
Post group work result – pre group work result	3.22	3.5011	.3083	10.449	.000

The result of the paired samples test at $\alpha = 0.05$ confirm that there is a significant difference between the pre- and post-group work exam results, favouring the post-group work exam results.

(ii) Hours of Attendance vs. Change in Performance

A regression analysis was also carried out in order to explain the relation between total hours of group work attendance and change in performance. The dependent variable, in this case, was the change in performance. The following is a summary generated by the SPSS package.

Table 7.3: Regression analysis[38] of hours of attendance and change in performance

Model	Un-standardized coefficients		Standardized coefficients		
	B	Std. error		Z	Sig
Constant	-.4.838	.744	Beta	-6.499	0.000
Total hours attended	.249	.022	0.709	11.324	0.000

This confirms that the number of hours of group work and change in performance are significantly related at $\alpha = 0.05$, i.e., students who attended group work for more hours performed significantly better than students who did not ($z = 11.324$, $p<0.05$).

The test for significance of correlations made at $\alpha = 0.05$ also showed that a significant correlations ($r = 0.709$) exist between hours of attendance in group work and change in performance. Furthermore, the coefficient of determination revealed that about 50% of the variation in change of performance was explained by total hours of attendance in the group work.

On the basis of this statistical evidence, one may conjecture that, over and above making a student join what seems a reasonably heterogeneous group, how much a student spends in group work significantly affects his/her performance.

[38] The output of the table is read as follows

- The **Unstandardized coefficients (B)** are the regression coefficients. The regression equation is
  change in performance= -4.838 + 0.249 hours.

- The **Standard Errors** are the standard errors of the regression coefficients
- The **Standardized coefficients (Beta)** are what the regression coefficients would be if the model were fitted to standardized data, that is, if from each observation we subtracted the sample mean and then divided by the sample SD.
- The **z** statistic tests the hypothesis that a population regression coefficient β is 0, that is, H_0: $\beta= 0$. It is the ratio of the sample regression coefficient β to its standard error
- **Sig.** labels the **two-sided P values** or **observed significance levels** for the t statistics

7.3.2 Comparison of Grouping Methods

The following tables exhibit a cross tabulation of the grouping methods by change in performance.

Table 7.4: Cross tabulation of grouping methods by change in performance

		Grouping Method						Total
		Program Assigned		Self Assigned		Randomly Assigned		
Change in performance	Decreased		0%	3	6%	4	10,53%	7
	No change	13	31.71%	23	46%	17	44,74%	53
	Increased	28	68.29%	24	48%	17	44,74%	69
Total		41		50		38		129

As may be observed from the table above, the program-assigned method has the highest proportion of students who have increased in performance (68.29%) followed by those who were in self-assigned groups.

In addition to what is revealed by the percentage figures, a statistical test was carried out in order to examine which grouping method is better in terms of yielding a higher proportion of increase in performance (referred to as success). The test used for this purpose was the two-sample test for proportion. For the purpose of carrying out the statistical test, proportion of success was defined as "the proportion of those who have increased their performance" and proportion of failure referred to "those who have not increased performance (those who have decreased or not changed their performance)". The following proportions of success and failure were summarized from Table 7.4.

Table 7.5: Proportions of success in the three grouping methods

	Grouping method		
Proportion	Program-assigned	Self-Assigned	Randomly-Assigned
No. of students	41	50	38
Success	0.683(pg)	0.480(sg)	0.447(rg)
Failure	0.317	0.520	0.003

- Note that "pg" refers to the proportion of success in program-assigned group, "sg" refers to the proportion of success in self-assigned groups and "rg" refers to the proportion of success in randomly-assigned groups. The capital letter PG, RG and SG refer to the population proportion.

The test statistic for the difference of proportion as presented in Gupta (1992) is

$$Z = \frac{p_1 - p_2}{\sqrt{PQ(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Where p_1 and p_2 refer to proportion of success with two different grouping methods,

n_1 and n_2 refer to the size of sample and

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

The following two tests were made based on the above formula.

Test of Significance for Difference of Proportions between Program-Assigned and Self - Assigned Methods

Here the null hypothesis was that the population proportion of students who have increased performance in both program-assigned and self-assigned grouping methods is same.

H_o: PG = SG.

The alternate hypothesis stated that the proportion of students who have increased in performance is significantly higher in program-assigned method than the self-assigned method.

H_1: PG > SG

The significance value of Z for one-tailed test at 5% level of significance is 1.645 while the calculated Z statistic reveals a value of 1.93. Since the calculated value is greater than 1.645, we can safely claim that there is a significant difference in academic performance between students who attended the two grouping methods and in favour of the Program-Assigned method.

Test of Significance for Difference of Proportions between Program-Assigned and Randomly-assigned Methods

The null hypothesis, in this case, was that the population proportion of students who have increased performance in both program-assigned and randomly-assigned grouping methods is same.

H_o: PG = RG.

The alternate hypothesis stated that the proportion of students who have increased in performance is significantly higher in program-assigned method than the randomly-assigned method.

H_1: PG > RG

Since the calculated Z statistic reveals a value of 2.165, we can safely conclude that there is a significant difference in academic performance between the two grouping methods and in favour of the program-assigned method.

From the results of the above two statistical tests, one can generally conclude that performance has definitely increased as a result of group work. What is more, the program-assigned method has significantly improved performance of students as compared to both the self-assignment and random-assignment methods.

In this connection, it is also interesting to note that, those who have above satisfactory performance before group work have not decreased their performance, rather their results either have improved or remained unchanged. Based on the results, we can safely claim that while low achievers improved their performance significantly, there is no loss of performance from high ability students – an observation in accord with Slavin (1990) and Kulik and Kulik (1992) who stated that ability grouping has essentially no negative effect on student achievement.

One may come out with various reasons why the students with program-assigned groups have performed better than the others. For instance, socialising, exaggerated funs, and private matters might not have been exercised since most students were grouped based on their characteristics. Moreover, we find that there is at least one motivated/serious student in the group who encourages the group work.

It has been noted, however, that there were regular absentees from group work. These might be attributed to their lack of willingness and seriousness. But we can not overlook the possibility that students might have been required to go straight home from school instead of staying for group work. To address this issue, there should be a mechanism to create awareness for parents as well as students on the benefits of group work.

Another issue worth raising is the topics that the students have been working on during group work. The study focused on specific topics that the students have been doing during the first semester of grade 11 (these are relatively easier topics as compared to those in the second semester). The topics selected might have positively contributed to the observed better level of performance of the students. As such, one might need to conduct the research further with more difficult topics (for instance Geometry topics) in order to ensure the consistency of the results.

In relation to the incremental version of the algorithm, a number of advantages may be gained since it does not require the whole student data records in advance. One may apply the incremental version to suggest to a student a group he/she might like to join. It could also be used in adaptive computerized tutoring systems (Intelligent Tutoring Systems) where individualized instruction is mostly practiced. This can be done by extending the function of the student model component of the intelligent tutoring system where in addition to storing the student's knowledge of the subject matter, personality attributes relevant for group composition might be stored. Suggestions may be also made by a software agent designed for the purpose. Such a direction is pointed out in Chapter 8 below.

7.4 Feedback from Students

At the end of the Group work, a survey was conducted to get feedback from students. A total of 140 students, who participated in the group work, were made to fill out a post-group work questionnaire. These were 43 from program-assigned groups, 52 from self-assigned groups, and 45 from randomly-assigned groups. In what follows, we present the summary of the findings from the questionnaires.

(i) Opinion on group formation

With regard to opinion of students on group formation, they were asked to comment on the size of the group and selection of group mates.

In response to the question related to the opinion on the size of the group which they have participated in, more than half of the students (63.6%) indicated that the group size was optimal. This opinion is highly shared specially with the program-assigned students (79.1%).

The results in preference on selection of group mates revealed rather confusing results. A higher proportion (65.1%) of those assigned with program preferred to be with their friends while higher proportions (65.5%) of those who have made self-assigned groups prefer other methods of grouping. The possible reasons for these may be that some students were not really serious at the time of filling out the questionnaire. The groups that were made based on self-selection might not have been really serious on the group work which also confirmed what is available in the literature on the disadvantages of self-assigned groups, i.e., they tend to socialize and talk more than studying. On the other hand, those students who were assigned with program may not have been happy since the group assignment has not been their own choice.

(ii) Opinion on how well they worked together

In order to get feed back on how well students worked together, they were asked to comment on the extent of participation during group work, the extent of efforts applied to do each question diligently, and willingness to work together again.

From the responses of students on the extent of participation during group work, a higher proportion of students in the program-assigned group (51.2%) indicated that all member of the group have participated during group work as compared to students in both self-assigned group (44.2%) and randomly-assigned group (31.1%).

It is interesting to note that higher proportion of students (more than half) who were assigned based on the program, felt that all have participated in the group work, while one might have expected this kind of information from the friends group who have made groups by choice. This is an observation partly supporting the idea that students with different behaviours (heterogeneous) groups might work more seriously than students assigned with other grouping methods.

In relation to the question of extent of efforts to do each question diligently, 59.3% of the students have indicated that they have made every effort to do each question accurately.

Interesting also is the fact that a higher percentage of students (64.4%), in randomly-assigned groups, indicated that they have made every effort to solve each question diligently. However, in the case of self-assigned groups, a smaller proportion of students have made efforts to solve questions diligently. One main reason for this may be, when sitting together as a group, students who are friends tend to socialize and discuss matters other than the subject under consideration. Of course, there were also students who were serious and motivated to indicate that their group was serious in solving the problems.

From the responses of students to the question of willingness to work with the same group in the future, although we expected higher proportion of the self-assigned groups to show positive answers, the findings were otherwise. A higher proportion of students (53.5%) in program-assigned groups said they were willing to work with the same group as compared to the students in either of the grouping methods, i.e., 40.5% in self-assigned and 31.1% in randomly-assigned.

(iii) Benefits from the Group Work

Students were also asked to comment on the benefits they gained from the group work; whether they feel the group work was successful; and what they thought were unfavourable conditions during group work.

It was observed from the responses that in all the grouping methods, higher proportion of students (65.0%) indicated that they improved their abilities to work with others. It is also seen that 46.4% of the students said they have learned other ways of solving problems which might have also contributed to their increase in performance.

The group work was beneficial to almost all the students. With the program-assigned group, a higher proportion(74.4%) agreed that the group work has improved their abilities to work with others. This also goes in conformity with the findings of other researchers (Johnson and Johnson,1986; O'Donnell and Dansereau,1992; and Bradley and Herbert,1997) who claimed that apart from improvement in performance, group work increases interaction abilities.

Most of the students felt that the group work was successful. It is interesting to note here that smaller proportion (44.2%) of students in self-assigned group felt that the group work was successful as compared to randomly-assigned students(51.2%) and program-assigned students (46.5%).

In relation to unfavourable conditions during group work, it was observed that a higher proportion of students (42.3%) in the self-assigned group were bored of regular meetings. One reason for this may be the strict schedule applied by the researcher in order to complete the experiment on time. As could be expected, some of the students in program-assigned (20.9%) and randomly-assigned (24.4%) groups indicated that they did not like the group they were assigned to.

Taken together, from this survey, we may generalize that students did favour group work and no significant negative impacts were observed.

7.5 Implications in the Field of Education

The results of this research work do have a number of implications in the field of education. For instance, the performance prediction model developed might help researchers in the field of education who are involved in identifying factors affecting performance. This aspect of the study may also be considered as an insight into the possible applications of uncertainty management techniques, particularly to address some of the conflicting results in relation to the significance of some factors affecting student performance. Moreover, researchers might even be able to extend the application of similar performance prediction models for other academic subjects.

In relation to the implications of the research to instructors, first the automatic tool developed may be of use to collect information from students and make heterogeneous groups with minimum or no manual intervention. From the relatively high accuracy of the performance prediction model, instructors might be able to use the model in cases where exam administration becomes difficult or impossible and where number marks are not required. What is more, the automated instrument to measure attributes might help Educational Psychologists to easily get information on the personal characteristics of a student and recognize areas where the student needs help. In this direction, however, one needs to be cautious of the privacy issues. Mechanisms need to be developed to protect misuse of the data obtained from a student.

The automated tool developed might also be of use to students. In a computerized environment, students might be free to measure their personal attributes confidently and with out the intervention of instructors or other individuals. They might also get information to help them discover their true self instead of mistaken self-concept. Students might have a good chance of knowing their predicted level of performance in mathematics and act accordingly.

In relation to the mathematics subject, very often, it is taught by a lecture-discussion format in large class sizes. The results of this study showed that group work might be incorporated in addition to lectures so that students interact, share ideas and study together to improve their performance in mathematics.

CHAPTER EIGHT

8. CONCLUSION AND DIRECTIONS FOR FUTURE WORK

8.1 Conclusion

As clearly indicated from the outset, this research generally aimed at exploring a computer-based approach for the purpose of forming effective heterogeneous groups by taking into account both the level of academic performance of students and personality attributes. In particular, the work involved the following three specific areas: (i) the identification and measurement of personality attributes, (ii) the development and evaluation of automated tools to predict the performance level of students and (iii) the development and evaluation of automated tools to form effective heterogeneous groups based on the results of (i) and (ii).

The attributes identified as relevant for consideration in group composition were also explored for applicability to predict performance. The Bayesian Performance prediction model developed as part of this study revealed information on how personality attributes identified to intervene with performance affect each other and on how questions of dependencies between some of the attributes can also be answered straight from the network.

In general, the results of the group composition experiment confirmed that group learning improves performance. The evaluation results indicated that students grouped based on level of performance and personality attributes, performed better than the randomly-assigned or self-selected groups and the developed automatic tool has proved to be a viable grouping technique to create effective groups.

It is interesting to note that the research looked into the different composition of personality traits and is able to present a logical deduction that diversity in the personality type will further enhance the performance of the group. While there are many ways to arrange students to work in cooperative/collaborative groups, the personality test and automatic grouping may be an option.

In view of the foregoing, we may conclude that cooperative learning methods hold great promise for accelerating interest of students and attainment of improved level of performance in mathematics.

8.2 Directions for Future Work

Three suggestions for future research work may emerge from the results of the experiment and the discussions that followed.

Inclusion of more attributes to improve the Performance Prediction model

The Bayesian network learning in the current work is based on eight attributes. In order to capture some other characteristics for consideration, further investigations may need to be done. Future lines of research might explore more deeply and more specifically the factors related to performance and for incorporation in the prediction model.

Improvements in the Automation Tool

The automation tool for group composition in this study has been developed with a mathematical model that gives equal weight to all the factors affecting performance. This is due to the lack of proper justification on which attributes are more important than the others. Even if the information on relevance is available, quantifying the weights requires more detailed examination of the attributes. In addition to detailed consultation with Educational Psychologists, this may require the use of such tools as decision trees to determine the information value of each attribute. As such, improving the grouping tool by revisiting the algorithm through the incorporation of such weights (that indicate the relative importance of the attributes) in the vector representation is worth exploring. Moreover, incorporating the mathematical model into such areas as optimization techniques and genetic algorithms might be useful to generate more optimized groups.

This research work, being an experiment over a fixed period of time, considered values of personality attributes as constant. However, with the exception of gender, the personality attributes used in the research are dynamic in nature the values of which may change over the course of the group work as new knowledge and experience is gained by students from group interactions. Therefore, one major future work may be to deal with the variation in student behaviour expected over time as a result of group interaction, maturity or other experiences. For instance a shy(introvert) student in the long run might become an extrovert one or a student who is below satisfactory in English might improve as a result of group interaction or other academic and social factors. The opposite might also be true with a chance of having a student whose personalities might deteriorate. Moreover, it should be noted that as values of the attributes change, there might be a need to reorganize groups to allow more interaction among students.

In this connection, two types of software programs might be designed.

(i) A Program which updates student information (Software Program – I)

This is a software program that may be designed to keep track of the student's change of behaviours. For instance the software may

- record the responses of the student to personal questions occasionally posed by the system;
- supervise the interactions that the student makes with other group members;
- present some scenarios to decide on change of behaviour;
- decide whether the student has changed behaviours and
- update its belief about the student.

(ii) A Program which reorganizes groups (Software Program – II)

This is another software program that may be designed to provide personalized assistance (software-agent) to advise a student on alternative group assignments. More specifically, the program interacts with the student model

- to check the student's level of knowledge of the subject matter;
- to check whether software program – I has updated the student model in relation to changes in values of personality attributes;
- to register how long a student has been in the group and history of interaction with group mates;
- to decide whether the student needs to change groups;
- to recommend alternative group assignments.

In order to address the above issues, one may extend the function of an existing student model component of an Intelligent Tutoring System(ITS). The student model, in addition to storing information about student's learning pace, misconceptions and weaknesses during the learning session, it may further be used to dynamically maintain students' personality attributes.

REFERENCES

Ademe M., and Gebre B. 1990. The impact of objective type tests on the learning process of high school mathematics. *The Ethiopian Journal of Education.* 11 (1): 28-49.

Akinkugbe, O. 1994. *Nigeria and education: The challengers ahead*. Lagos: Spectrum Books Limited.

Allen, W. H., and Van Sickle, R. L. 1984. Learning teams and low achievers. *Social Education*, 48: 60-64.

Ames, C., and Ames, R. (eds.). 1985. *Research on motivation in education*. Orlando: Academic Press, Inc.

Anthony, W. S. 1973. The Development of extraversion, of ability and of the relation between them. *British Journal of Educational Psychology*. 43: 223-227.

Apple, D., et al. 2001. *A Handbook on cooperative learning. 3rd ed.* Illinois: Pacific Crest.

Aronson, E. 2000. *The Jigsaw classroom*. available from: http://www.jigsaw.org. Internet: accessed 6 October, 2005.

Atkinson, E. 2000. An investigation into the relationship between teacher motivation and pupil motivation. *Educational Psychology*. 20(1): 45-57.

Bandura, A. 1977. *Social learning theory*. Englewood Cliffs, New Jersey: Prentice Hall

Banks, J. H. 1964. *Learning and teaching arithmetic,* 2nd ed. Boston: Allyn and Bacon.

Baris-Sanders, M. 1997. Cooperative education : lessons from Japan. *Phi Delta Kappan V.* 78(8):619-623.

Baumrind, D. 1991. Parenting styles and adolescent development. In *The encyclopaedia on adolescence,* ed. J. Brooks-Gunn, R. Lerner and A.C. Petersen, 746-758. New York: Garland.

Bayes, T. 1763. An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society,* pp. 370–418.

Beckman, M. 1990. Collaborative learning: Preparation for the workplace and democracy. *College Teaching. 38(4):* 128-133.

Bradley, J. H., and Herbert, F. J. 1997. The effect of personality type on team performance. *Journal of Management Development*. 16: 337-353.

Bryan, L. H. 1996. Cooperative writing groups in community college. *Journal of Adolescent and Adult Literacy*. 40(3):.

Bull, S. 1998. Do it yourself models of collaborative student modeling and peer interaction, ed. B.P.Gottel, et al. *ITS'98 LCNCS*. 1452: 176-185.

Caplan, S., et al. 2002. Socio emotional factor contributing to adjustment among early entrance to college students. *Gifted Child Quarterly* . 46(2): 124-134.

Carron, G., and Chau, T. G. 1996. *The quality of primary schools in different development contexts*. Paris: UNESCO.

Chan, T. W., et al. 1995. Reciprocal-tutoring-kids: tutor tutee role playing systems. *The Proceedings of Artificial Intelligence in Education,* Washington, D.C. 226-233.

Cheng, J., and Greiner, R. 2001. *Learning Bayesian belief network classifiers: Algorithms and system.* In Proceedings of the Fourteenth Canadian Conference on Artificial Intelligence (AI).

Cheng, J., et al. 1997. An algorithm for Bayesian belief network construction from data. Proc. *6th International Workshop on Artificial Intelligence and Statistics,* Florida. 83-90.

Chickering, A. W., and Gamson, Z. F (eds.). 1991. Applying the seven principles for good practice in undergraduate education. *New Directions forTeaching and Learning*. no.47. San Francisco: JosseyBass.

Conati, C., and VanLehn, K. 1996. A student modeling framework for probabilistic on-line assessment of problem solving performance. *Proc. 5th International Conference on User Modeling,* Kailua-Kona, HI: User Modeling Inc. 75-82.

Christison, M. 1994. Cooperative learning in the EFL classroom. In *Teacher development: Making the right moves.* ed. T. Kral. *Selected Articles from the English Teaching Forum 1989-1993.* Washington: English Language Program Division.

Cohen, E. G. 1994. *Designing group work: Strategies for the heterogeneous classroom*. New York: Teachers College Press.

Cochran, W. G. 1977. *Sampling techniques. 3rd ed.* John Wiley and Sons: New York.

Collier, K. G. 1980. Peer-group learning in higher education: The development of higher-order skills. *Studies in Higher Education.* 5(1): 55-62.

Collins, A., and Brown, J. S. 1988. The computer as a tool for learning through reflection. In *Learning issues for intelligent tutoring systems,* ed. H. Mandl and A. Lesgold, 1-18. New York: Springer Verlag.

Cooper, G. F., and Herskovits, E. A. 1992. Bayesian method for the induction of probabilistic networks from data, *Machine Learning*. 9: 309-347.

Cooper, J. 1990. Cooperative learning and college teaching: Tips from the trenches. *Teaching Professor*. 4(5): 1-2.

Comelius, ML., and Cockburn, D. 1978. Influence of pupils' performance. *Educational Research.* 21(1):48-53.

Corno, L. 1986. The meta-cognitive control components of self-regulated learning. *Contemporary Educational Psychology*. 11: 333-346.

Cowie, H., et al. 1994. *Cooperation in the multi ethnic classroom: The impact of cooperative group work on social relationships in middle schools.* London: David Fulton Publishers.

Cowell, R. 1999. Introduction to inference in Bayesian networks. In *Learning in graphical models,* ed. M. Jordon, 9-26. Cambridge, Massachusetts: MIT Press.

Cox, D. R. 1967. The Adolescent in Ethiopia. *Ethiopian Journal of Education*. 1(1): 50- 56.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometica*. 16: 297-234.

Cronbach, L. J. 1970. *Elements of psychological testing.* N.Y:Harper.

Dansereau, D., and Johnson, D. 1994. Cooperative learning. In *Learning, Remembering, Believing: Enhancing Human Performance.* 83-111.

Davis, B. G. 1993. *Tools for teaching*. Jossey-Bass Publishers: San Francisco.

Davis, J. A. 1971. Use of measurement in student planning and guidance. In *Educational Measurement. 2nd ed,* ed. R.L. Thorndike, 671-679. Washington, D.C.: American Council on Education.

Dees, R. 1983. Cooperation in the mathematics classroom. In *Cooperative Learning in Mathematics: A Handbook for Teachers*. Edited by Neil Davidson. Menlo Park, CA: Addison-Wesley Publishing Company.

Diaz, A. L. 2003. Personal, family, and academic factors affecting low achievement in secondary school. *Electronic Journal of Research in Educational Psychology and Psycho Pedagogy*. 1(1): 43-66.

Dishon, D., and O'Leary, P. 1984. *A guidebook for cooperative learning: A technique for creating more effective schools*. Holmes Beach, FL: Learning Publications.

Druzdzel, M. J., and Diez, F. J. 2003. Combining knowledge from different sources in causal probabilistic models. *Journal of Machine Learning Research*. 4: 295-316.

Entwistle, N. J. 1972. Personality and academic attainment. *British Journal of Educational Psychology.* 42: 137-151.

Ernest, J. 1976. *Mathematics and sex*. Santa Barbara, CA : University of California.

Fenneman, E. 1974. Sex differences in mathematics learning. *Elementary School Journal*. 75(3): 183-190.

Florea, A. M. 1999. An Agent-based collaborative learning system: Advanced research in computers and communications in education, ed. G. Cumming et al, IOS press. *International Conference on Computers in Education'99*. 161-164.

Fuller, B. 1987. What factors raise achievement in Third World? *Review of Educational Research*. 57(3).

Fung, R., and del Favero, B. 1994. *Backward simulation in Bayesian networks*. In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann, San Francisco, CA, 227-234.

Fung, R., and Chang, K. 1990. *Weighting and integrating evidence for stochastic simulation in Bayesian networks*. In Henrion, M., Shachter, R.D., Kanal, L.N. & Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence 5. Elsevier Science Publishers B.V. (North Holland): 209-219.

Gall, A., et al. 1973. *Present problems in the democratization of secondary and higher education*. Paris: UNESCO.

Garden, R. A. 1987. The second IEA mathematics study. *Comparative Education Review.* 31: 47-68.

Gill, I. 1991. Does the structure of production affect demand for schooling in Peru? women's work, education and family welfare in Peru. *World Bank Discussion Paper* No. 16, Washington D.C: World Bank.

Girma G. 2003. In-service English Language Teacher's Perceptions of the Factors that Influence the Implementation of Group Work Activities in ELT Classes. *Ethiopian Journal of education* 23(2): 103-126.

Good, T. L. and Brophy, J. E. 1990. *Educational psychology: A realistic approach.* 4th ed. New York: Long Man.

Goodsell, A. M., et al. 1992. *Collaborative learning: A sourcebook for higher education.* University Park: National Center on Postsecondary Teaching, Learning, and Assessment, Pennsylvania State University.

Guilford, J. P. 1956. *Fundamental statistics in psychology and education*. New York: McGraw-Hill Book Company.

Gupta, S. C. 1992. *Fundamentals of statistics,* 5th ed. Himalaya Publishing House: Hyderabad.

Haile W. 1970. Social dependency in Ethiopia and its consequences of learning. *Ethiopian Journal of Education*. 4(1): 4-15.

Hake, R. R. 1998. Interactive-engagement vs. traditional methods: A six-thousand- student survey of mechanics test data for introductory physics courses. *American Journal of Physics*. 66(1): 64-80.

Heckerman, D. 1995. A tutorial on learning Bayesian networks. *Technical Report MSR-TR-95-06, Microsoft Research.*

Heckerman, D. 1999. A Tutorial on learning with Bayesian networks. In (Ed.), *Learning in graphical models,* ed. M. Jordon, 301-354. Cambridge, Massachusetts: MIT Press.

Henrion, M. 1988. *Propagating uncertainty in Bayesian networks by probabilistic logic sampling*. In Lemmer, J.F. and Kanal, L.N. (eds.) Uncertainty in Artificial Intelligence 2. Elsevier Science Publishers B.V. (North Holland), pages 149-163.

Holt, D. D. 1993. Cooperative learning for students from diverse language backgrounds: An Introduction. In *Cooperative learning: A response to linguistic and cultural diversity*, ed. Daniel D. Holt. McHenry, 1-8. Washington, D.C: Delta Systems and Center for Applied Linguistics.

Humphreys, B., et al. 1982. Effects of cooperative, competitive, and individualistic learning on students' achievement in science class. *Journal of Research in Science Teaching*. 19(5): 351-356.

Jacobs, G. 1988. Cooperative goal structure: A way to improve group activities. *ELT Journal*. 42 (2): 97-100.

Jensen, F. V. 1996. *An introduction to Bayesian network*. London. U.K: University College London Press.

Jensen, F. V. 1996. Bayesian network basics. *AISB Quarterly*. 94: 9-22.

Jensen, F. V., et al. 1990. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*. 4: 269-282.

Johnson, D. W., and Ahlgren, A. 1976. Relationship between student attitudes about cooperation and competition and attitudes toward schooling. *Journal of Educational Psychology*. 68(1): 92-102.

Johnson, D. W., et al. 1981. Effects of cooperative, competitive, and individualistic goal structures on achievement: A meta analysis. *Psychological Bulletin*, 89: 47-62.

Johnson, D. W., et al. 1986. *Circles of learning: Cooperation in the classroom*. Edina, MN: Interaction Book Company

Johnson, D. W., and Johnson, R. T. 1985. The internal dynamics of cooperative learning groups. In *Learning to cooperate, cooperating to learn, ed.* R. Slavin, et al., 103-124. New York: Plenum.

Johnson, D. W., and Johnson, R. T. 1986. Action research: Cooperative learning in the science classroom. *Science and Children*, October, 1986, pp. 31-32.

Johnson, D. W., and Johnson, R. T. 1989. *Leading the cooperative school*. Edina, MN: Interaction.

Johnson, D. W., and Johnson, R. T. 1990a. What is cooperative learning? In p*erspectives on small group learning: Theory and practice, ed.* M. Brubacher. Pmtarop: Rubican Publishing Ind.

Johnson, D. W., and Johnson, R. T. 1990b. Cooperative classrooms. In *Perspectives on small group learning: Theory and practice*. ed. M. Brubacher. Ontario: Rubican Publishing Ind.

Johnson, D. W., and Johnson, R. T. 1991. *Learning together and alone: Cooperative, competitive, and individualistic.* 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Johnson, D. W., and Johnson, R. T. 1994. *Learning together and alone: Cooperative, competitive, and individualistic learning.* 4th ed. Edina, Minn.: Interaction Book Company.

Jordan, D., and Le Metais, J. 1997. Emotional intelligence and student behaviour. *International Electronic Journal For Leadership In Learning* 1(2).

Kagan, S. 1994. *Cooperative Learning*. San Clemente, CA: Kagan Publishers.

Kamps, D. M., et al. 1994. Enhanced small group instruction using choral responding and student interaction for children with autism and developmental disabilities. *American Journal on Mental Retardation*. 99: 60-73.

Khandker, R. S. 1996. *Education, achievement and school efficiency in rural Bangladesh*. Washington D.C. World Bank.

Krause, P. J. 1999. Learning probabilistic networks*: The Knowledge Engineering Review.* 13(4): 321-351.

Krejins, K., et al. 2002. The sociability of computer-supported collaborative learning environments. *Educational Technology and Society.* 5(1): 26-37.

Kulik, J. A., and Kulik, C. C. 1992. Meta-analytic findings on grouping programs. *Gifted Child Quarterly*. 36(2): 73-77.

Lauritzen, S., and Spiegelhalter, D. 1988. Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society Series B*. 50:157-224.

Maccoby, E. E., and Jacklin, C. N. 1981. Myth, reality, and shades of gray: What we know and don't know about sex differences. In *Contemporary issues in educational*

Psychology 4th ed., ed. H.F. Clarizio, R.C. Craig, And W.A. Mehrens, 78-83. Boston: Allyn and Bacon.

Magland, N. E. A. 1994. School supply, family background and gender specific school enrollment in Sudan. *East African Social Science Research Review.* 10(2):1-20.

Martin, E., and Paredes, P. 2004. *Using learning styles for dynamic group formation in adaptive collaborative hypermedia systems*. available from http://www.ii.uam.es/~rcarro/AHCW04/MartinParedes.pdf. Internet: accessed 29 September 2005.

McConnell, D. 1994. *Implementing computer supported cooperative learning*. London: Kogan Page.

McGroarty, M. 1991. Cooperative learning and second language acquisition. In *Cooperative learning: A response to linguistic and cultural diversity*, ed. Daniel D. Holt. McHenry, 19-46. III and Washington, D.C: Delta Systems and Center for Applied Linguistics.

McKeachie, W. J. 1986. *Teaching tips: A guidebook for the beginning college teacher*. Lexington, MA: D.C. Heath and Company.

Mills, C. J. 1993. Gender differences in academically talented young students mathematical reasoning: Patterns across age and sub skills. *Journal of Educational Psychology*. 85: 340-346.

Murray, W. R. 1998. Practical approach to Bayesian student modeling, Proc. *4th International Conference on Intelligent Tutoring System.* 424-433.

Neale, D. C. 1969. The role of attitudes in learning mathematics. *The Arithmetic Teacher*. 16: 631-640.

Neapolitan, R. E. 2004. *Learning Bayesian networks*. Chicago, North Eastern: Illinois University.

Nijstad, B. A., and Carsten, K. W. D. 2002. Creativity and group innovation. *Applied Psychology*. 51(3): 400.

Nunnaly, J. 1978. *Psychometric theory*. New York: McGraw-Hill.

Oakes, J. 1990. *Multiplying inequalities: The effects of races, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: RAND.

O'Donnell, A. M. and Dansereau, D. F. 1992. Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In, *Interaction in cooperative groups: The theoretical anatomy of group learning*, ed. R. Hertz-Lazarowitz and N. Miller, 120-144. New York: Cambridge University Press.

Ormord, J. E. 2003. *Human learning.* 4th ed. Englewood Cliffs, NJ: Merril Prentice Hall.

Panitz, T. 1999. Using cooperative learning 100% of the time in mathematics classes establishes a student-centered, interactive learning environment. available from: http://home.capecod.net/~tpanitz/tedsarticles/coopmath.htm. Internet: accessed 6 October, 2005.

Pearl, J. 1986. *Fusion, propagation, and structuring in belief networks*. Artificial Intelligence, 29(3): 241-288.

Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann: San Mateo CA.

Perreault, R. J. 1983. An experimental comparison of cooperative learning to non-cooperative learning and their effects on cognitive achievement in junior high industrial arts laboratories. (Doctoral dissertation, University of Maryland, 1982). *Dissertation Abstracts International*, 43(3830A).

Raph, J. B., et al . 1966. *Bright underachievers*. New York, Teacher College Press.

Romney, C. 1996. *The benefits of collaborative learning*. Available from http://www.ucalgary.ca/pubs/Newsletters/Currents/Vol3.6/Benefits.html. Internet: accessed 20 August 2005.

Rose, P., et al . 1997. *Gender and primary schooling in Ethiopia*. Brighten: IDS.

Rothstein, P. R. 1990. *Educational Psychology*. New York: McGraw-Hill.

Russel, S., and Norvig, P. 2003. *Artificial intelligence: a modern approach*, 2nd ed. NJ: Prentice Hall.

Schiefelbein, E., et al. 1994. Repeating: an overlooked problem of Latin American education. *Comparative Education Review.*19: 468-87.

Shachter, R. D., and Mark, A. P. 1990. *Simulation approaches to general probabilistic inference on belief networks*. In Henrion, M., Shachter, R.D., Kanal, L.N. & Lemmer, J.F. (eds.) Uncertainty in Artificial Intelligence 5. Elsevier Science Publishers B.V. (North Holland), pages 221-231.

Sherman, J. 1980. Mathematics, spatial visualization, and related factors: changes in girls and boys, grades 8-11. *Journal of Educational psychology*. 72: 476-482.

Slavin, R. E. 1983a. When does cooperative learning increase achievement? *Psychological Bulletin*. 94: 429-445.

Slavin, R. E. 1983b. Student team learning in math. In *Cooperative Learning in Mathematics: A Handbook for Teachers*. Edited by Neil Davidson. Menlo Park, CA: Addison-Wesley Publishing Company.

Slavin, R. E. 1987. Developmental and motivational perspectives on cooperative learning: A reconciliation. *Child Development*. 58: 1161-1167.

Slavin, R. E. 1990. Achievement effects of ability grouping in secondary schools a best-evidence synthesis. *Review of Educational Research*. 60(3): 471-499.

Slavin, R. E. 1991. *Educational psychology.* 3rd ed. Englewood Cliffs, NJ: Prentice Hall.

Slavin, R. E. 1995. *Cooperative learning: Theory, research, and Practice,* 2nd ed. Boston: Allyn and Bacon.

Slavin, R. E. 1996. *Every child, every school: Success for all.* Thousand Oaks, CA: Corwin Press.

Spiegelhalter, D.J., et al. 1993. Bayesian analysis in expert systems. *Statistical Science* 8: 219-247.

Spirtes, P., et al. 1993. Causation, prediction, and search. *Springer-Verlag Lecture Notes in Statistics 81***.** New York: Springer-Verlag.

Stepaneck, J. 1999. *The inclusive classroom, meeting the needs of gifted students: differentiating mathematics and science instruction.* Northwest Regional Educational Laboratory.

Stromquist, N. 1997. *Literacy for citizenship: gender and grassroots dynamics in Brazil.* Albany, NY: SUNY Press.

Swain, M. 1985. Communicative competence: some roles of comprehensive output in its development, in *Input in second language acquisition*. ed. Susan M. Gass and Carolyn G. Madden, 235-253. Boston, Mass.: Heinle and Heinl.

VanLehn, K., and Martin, J. 1995. Student assessment using Bayesian nets. *International Journal of Human Computer Studies . 42*:575-591.

VanLehn, K., et al. 2002. Minimally invasive tutoring of complex physics problem solving. Proc. *5th International Conf. on Intelligent Tutoring Systems*: 367-376.

Webb, N. M. 1992. Testing a theoretical model of student interaction and learning in small groups. In *Interaction in cooperative groups: a theoretical anatomy of group learning,* ed. R. Hertz-Lazarowitz and N. Miller, 102-119. Cambridge: University Press.

Whilhite, S. C. 1990. Self-efficacy, locus of control, self-assessment of memory ability, and study activities as predictors of college course achievement. *Journal of Educational Psychology.* 82(4): 696-700.

Wilkons, L., and Marrette, C. 1985. *Gender influences in classroom instruction.* Orlando, FL: Academic Press.

Williams, M., and Burden, R. 1997. *Psychology for language teachers: a social constructivist approach*. Cambridge: Cambridge University Press.

Wodarski, L. A., et al. 1980. Teaching nutrition by teams-games-tournaments. *Journal of Nutrition Education*, 12(2): 61-65.

Zimmerman, B. J., and Martinez-Pons, M. 1986. Development of structured interview for assessing student's use of self-regulated learning strategies. American Educational Research Journal. 23: 614-628.

Zimmerman, B. J., and Martinez-Pons, M. 1988. Construct validation of a strategy model of student self-regulated learning. *Journal of educational Psychology*. 80: 284-290.

Zsolnai, A. 2002. Relationship between children's social competence, learning motivation and school achievement. *Educational Psychology.* 22(3): 317-330.

Unpublished References

Adane T. 1993. *Wastage in primary schools of Bahirdar Awraja*. M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Bedilu W. 1996. *The Correlation between extrovert-introvert personalities and ability to read and write essays*. M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Birhanu A. 1996. *The relationship of parenting styles with academic achievement among senior secondary school students: with particular reference to the kaffecho zone.* M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Daniel T. 1992. *Personality Types and academic achievement motivation as correlates of college academic performance*. M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Darge W. 1989. *The Reactions of social sciences first year students in Addis Ababa University to moral dilemmas related to academic matters.* Addis Ababa University, Addis Ababa.

Demis Z. 1991. *Factors that influence academic performance of trainees in the Harar Teacher Training Institute.* M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

High, C. F. 1996. A regression analysis of selected factors that influence the scores of students on the TASP Test. A Paper Presented at *the Annual Conference of the Texas Association of College Test Personnel, Texas*.

Markos T. 1996. *The relationship between parenting style and school performance among high school students in Mekelle.* M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Mathewos T. 1995. *A Study on factors affecting academic achievement of students at Bole Senior Secondary School*. M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Mekonnen Y., and Darge W. 1991. *The comparability of the predictive power of the Ethiopian School Leaving Certificate Examination with other measures* “preliminary observation”. Paper presented at the 1990/91 annual seminar of the Institute of Educational Research, Addis Ababa University.

Rahel B. 2002. *Reflections of freshman students towards learning mathematics,* Addis Ababa University, Addis Ababa.

Sewnet M. 1995. *Some factors affecting scholastic achievement of elementary school pupils.* M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Taddese T. 1990. *The Interrelations of English language proficiency and academic achievement of seventh grade Ethiopian students.* M.A. Thesis, School of Graduates Studies. Addis Ababa University, Addis Ababa.

Tsigie H. 1991. *An Assessment of the academic performance of female students in higher education institution in Ethiopia*. Addis Ababa University, Addis Ababa.

Yelfign W. et al. 1995. *Primary school female Participation and Performance in Checha District.* Ministry of Education. Addis Ababa.

Zaudneh Y., Darge W., and Nardos A. 1989. *A survey of the teaching – learning situation in institutions of higher learning in Ethiopia.* Addis Ababa University, Addis Ababa.

APPENDICES

APPENDIX – A

Survey Questionnaire to identify relevant attributes for group formation and performance prediction

This survey questionnaire is prepared in order to validate the relevant attributes known to affect the success of group work as well as performance in mathematics.

Please indicate the extent of your agreement (by putting an "X" mark) whether to consider the attribute in group formation and whether the same attributes would also be a factor for determining level of performance.

Thank you.

		Should be considered In group formation			**It can be considered as factor for performance**		
No.	**Attributes**	**Strongly Disagree**	**Undecided**	**Strongly Agree**	**Strongly Disagree**	**Undecided**	**Strongly Agree**
1	Mathematics Performance						
2	Seriousness/Dedication						
3	Gender						
4	English language ability(English performance)						
5	Family educational background						
6.	Family socio economic background						
7.	Ethnic background						
8.	Group Work Attitude						
9	Age						
10	Shyness(introvert personality)						
11	Religion						
12.	Interest for Mathematics						
13	Achievement motivation						
14.	Self confidence (Internal Locus of Control)						

APPENDIX - B

Judges who participated in the evaluation(rating) of the items prepared to measure the attributes.

Position	**Sex**	**Qualification**	**Specialty**
Associate Professor	M	Ph.D.	Psychology
Associate Professor	M	Ph.D	Psychology
Assistant Professor	M	M.A.	Psychology
Assistant Professor	F	Ph.D.	Psychology
Lecturer	M	M.A.	Psychology
Lecturer	M	M.A.	Psychology
Lecturer	M	M.A.	Psychology
Lecturer	M	M.Sc	Business Education and Information Science
Lecturer	M	M.Sc	Mathematics and Information science
Lecturer	M	M.Sc	Mathematics

APPENDIX - C
Instrument developed to measure attributes: English version

Dear Student
Each of the following statements expresses a feeling or attitude. There are no right or wrong answers. It is simply a measure of your opinion regarding the behaviour explained in each item.

Please indicate the extent of your agreement by putting an “X” mark for each statement, on the basis of your own true personal feelings. Answer all the items quickly and do not think too long about the exact meaning of each question.

For the researcher’s convenience in analyzing the data, you are kindly requested to write your Roll No. Please be assured that your Roll No. or full name will not appear in the actual research work, and all the data collected will be destroyed once the experiment is completed.

Thank you very much

1. Roll No. ___________________ 2. Section ____________________

3. Gender _______Male
_______Female

4. Highest Level of Education completed by your parents or guardians

Father’s (Male Guardian)

_______Illiterate _______Primary _______ Secondary

_______Diploma _______First Degree _______Second Degree and Above

5. Mother’s (Female Guardian)
_______Illiterate _______Primary _______ Secondary

_______Diploma _______First Degree _______Second Degree and Above

		Strongly Agree	Agree to some extent	Strongly Disagree
1.	I feel responsible when I am assigned to study with other students.			
2.	I feel comfortable working with my friends.			
3.	I understand the subject better when I explain the method of solving problems to my fellow students.			
4.	I learn variety of approaches for solving a problem when I study in groups.			
5.	Studying in groups give me the opportunity to discuss and clarify ideas.			
6.	Groups help me improve communication and social skills.			
7.	I never told lies.[39]			
8.	I find study groups as more enjoyable learning environments than conventional lectures.			
9.	I get more confidence when I study with my friends than alone.			
10.	I spend a lot of time when I study in groups.			
11.	Students are not often serious when they study in groups.			
12.	Group work is fun.			
13.	I am always under a terrible strain in a group work.			
14.	I never get tired of studying in groups.			
15.	.Studying in groups makes me stimulated.			
16	I never bothered about my exam results.[39]			
17	Studying in groups makes me feel secure.			
18.	Mathematics is my favourite subject.			
19.	If someone suggested that I take up maths class as my life's work, I would reply YES.			
20.	I intend to take other mathematics courses if I get the opportunity.			
21.	I never get tired of solving new problems in mathematics.			
22.	I fail in mathematics because I lack the ability .			
23.	I like challenging mathematics questions.			
24.	I never picked something which I found accidentally.[39]			
25.	I am curious on solutions of maths problems.			
26.	I do not like to do mathematics in my free time.			
27.	My mathematics class is boring.			
28.	My mathematics class has enjoyable assignments.			
29.	I have usually clear idea of what the mathematics class is all about.			
30.	I feel confident of solving maths problems as a result of my background in mathematics.			
31.	Learning mathematics sharpened my analytic skills.			
32.	I want to take other mathematics courses if I get the opportunity.			
33.	I am satisfied with my overall performance in maths.			
34.	I never envied my friends or other people.[39]			
35.	I am very much concerned with my results and I do not want to miss what the teacher teaches in class.			
36.	I must always get prizes and grades.			
37.	I use my abilities to the maximum to study my lessons.			
38.	When I suspect that there are going to be questions on an examination from outside reading assignments, I always read related materials.			
39.	My interest in class and subsequent academic achievements is better than all the other students.			
40.	I failed today does not mean that I will fail next time.			
41.	I must get college admittance so as to make my parents expectation fulfilled.			
42.	When I begin to do something, I do not get rest until its successful completion.			

[39] A lie detector statement

		Strongly Agree	Agree to some extent	Strongly Disagree
43.	I never touch my friends' belongings with out their approval. [39]			
44.	when I get less grades in my exams, I feel frustrated and upset to the extent that I do not want to eat.			
45.	I stick to my home work until it is completed even if it gets late.			
46.	I must study even during my free time.			
47.	When I'm given challenging questions, I sit through the night trying to get a solution to it.			
48.	I usually tackle the easy problem first and do not worry about the more difficult ones.			
49.	I reach to the extent of hating myself, if do not finish my work successfully.			
50.	I must excel academically from my classmates.			
51.	It is easy for me to speak up in class.			
52.	Teasing does not necessarily mean that peers do not like me.			
53.	I never fought with anyone.[39]			
54.	I feel satisfied of myself when I solve difficult problems.			
55.	I know and defend my stand when I am with other people.			
56.	I do not really believe in luck or chance.			
57.	I am not ashamed of the wrong answers I give in class, rather I tend to correct them.			
58.	I earn the respect and honours I receive.			
59.	I am not easily hurt when people find fault with me or the work I do.			
60.	When I make plans, I am almost certain that I can make them work.			
61.	There is a direct connection between how hard I study and the grades I get.			
62.	I get less grades in my lessons since I do not study hard.			
63.	Sometimes I can't understand how teachers arrive at the grades they give.			
64.	I am not easily disturbed with what people say as long as what I think is right.			
65.	I believe in my decisions and are by no means influences from my friends.			
66.	I often withhold my opinions for fear that It might be wrong.			
67.	I feel tense when I am with people I don't know well.			
68.	I have never gone against my parents' advice.[39]			
69.	I do not have trouble making friends at first.			
70.	I am not afraid of reading out loud in front of others.			
71.	I get nervous when I am asked to speak up in class.			
72.	I have very good abilities in my social Interactions.			
73.	I feel embarrassed when people complement me.			
74.	I prefer asking my friends later instead of the instructor whenever I have questions.			
75.	It is easy for me to be familiar with my classmates in a short time.			
76.	I do not get it difficult mixing with other students when I change my school.			
77.	I never had a teacher whom I disliked. [39]			
78.	I am apprehensive about going into a room full of students I do not know.			
79.	I easily make friends with other students.			
80.	I may be classified as one of the outspoken students in class.			
81.	I am shy of speaking first when I meet new people.			
82.	I spontaneously introduce myself to new students.			
83.	I prefer to keep quite in get together parties.			

APPENDIX - D
Mathematics exam administrated to students
(Pre-group work)

Name: ____ section ________

1. **Simplify**

$$3\left(\frac{-2}{5}\right)\left(\frac{-1}{4}\right)+\frac{5}{8}\left(\frac{-2}{3}+6\right)\div\frac{-16}{3}$$

2. Give additive inverse of $(-a+\frac{2}{3})$

3. Give the multiplicative inverse of $(-a+\frac{2}{3})$ (where a $\neq \frac{2}{3}$)

4. Find the greatest common divisor of 252 and 294

5. Simplify

$$\left(\frac{81x^{-2}y^{6}}{16x^{6}y^{-10}}\right)^{3/4}$$

6. Simplify by substituting $x=\frac{1}{2}$, and $y=\frac{3}{4}$

$$\frac{x^{-1}y-2x^{-2}y^{-1}}{3xy^{-2}-(2y)^{-1}}$$

7. Simplify $\sqrt[3]{-81}-4\sqrt[3]{-192}-\sqrt[3]{-375}$

8. If a = 2 and b = 1 simplify

$$\left(\frac{a^{-3}}{b^{4}}\right)^{\frac{1}{12}}-\sqrt{a^{3}}+\sqrt[3]{-81}$$

9. Simplify

$$(5t^{5}-2y^{2}+3y-4)-(3y^{3}+4t^{5}-y^{2}+7)$$

Solve for x, for each of the questions below (question 10-12),

10. 2(x – 1) + 7 (2x + 3) = 16

11. $2-\frac{3x}{8}\leq\frac{5x}{2}+\frac{3}{4}$

12. $-(3x+7)\leq 8-2x\leq(3x+7)$

13. Simplify

$$\frac{x^{-1}y-2x^{-2}y^{-1}}{3xy^{-2}-(2y)^{-1}}$$

14. $\left(\frac{8a^{-4}b^{e}}{27a^{2}b^{-3}}\right)^{1/3}$

15. If X = 3, Y =2, Simplify the result of $\frac{2x^{-1}-3y^{-1}}{x^{-2}-y^{-2}}$

APPENDIX – E
Experts considered and Survey content to solicit expert opinion on modification of the Bayesian network

Experts considered in the Survey

Experts	Considered as potential participants	Actually participated[40]
High school mathematics instructors (all male)	15	12
Freshman mathematics teachers (all male)	15	11
Educational Psychologists (all male)	15	15
English Teacher (all male)	4	3
Women instructors - (Gender experts, Freshman mathematics instructors and Educational psychologists)	7	6
	56	47

Survey Content

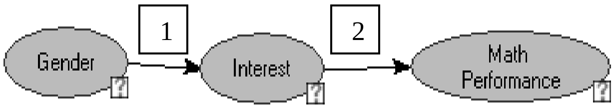
I. Please tick in the boxes to show your agreement on the statements provided.

No	**Dependencies**	**Response**
1	Attitude affects confidence	___ Yes ____ No
2	Confidence affects attitude	___ Yes ____ No
3	Confidence affects extent of mathematics performance	___ Yes ____ No
4	Confidence affects motivation	___ Yes ____ No
5	Confidence affects shyness	___ Yes ____ No
6	English ability affects extent of mathematics performance	___ Yes ____ No
7	English ability affects group work attitude	___ Yes ____ No
8	gender affects extent of shyness	___ Yes ____ No
9	Gender affects extent of mathematics performance	___ Yes ____ No
10	Gender affects interest for math	___ Yes ____ No
11	Interest for mathematics affects confidence	___ Yes ____ No
12	Interest affects extent of mathematics performance	___ Yes ____ No
13	Mathematics performance affects extent of confidence	___ Yes ____ No
14	Mathematics performance affects gender	___ Yes ____ No
15	Mathematics performance affects extent of English performance	___ Yes ____ No
16	Mathematics performance affects group work attitude	___ Yes ____ No
17	Mathematics performance affects interest for math	___ Yes ____ No
18	Achievement motivation affects Attitude towards group work	___ Yes ____ No
19	Achievement motivation affects confidence	___ Yes ____ No
20	Achievement motivation affects mathematics performance	___ Yes ____ No
21	Shyness affects extent of confidence	___ Yes ____ No
22	Shyness affects extent of English performance	___ Yes ____ No
23	Shyness affects gender	___ Yes ____ No
24	Shyness affects group work attitude	___ Yes ____ No

[40] The distributed and actually considered are different since some did not return the survey questions.

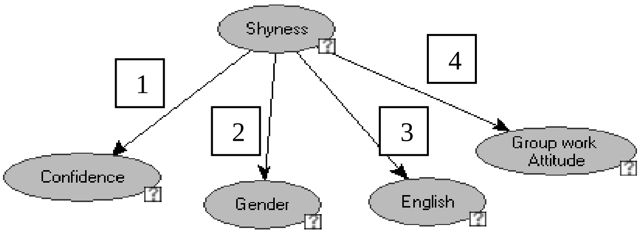
I. Please state your agreement and disagreement for the following causal diagrams. (Arrows indicate cause)

(i)



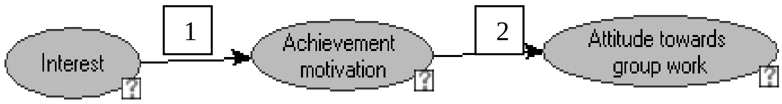
1 ___ Yes _____ No
2 ___ Yes _____ No

(ii)



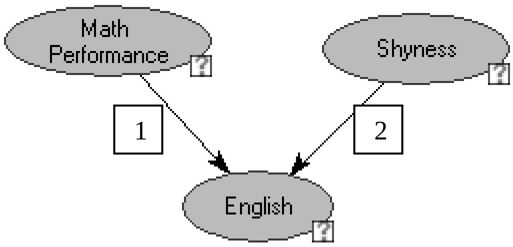
1 ___ Yes _____ No
2 ___ Yes _____ No
3 ___ Yes _____ No
4 ___ Yes _____ No

(iii)



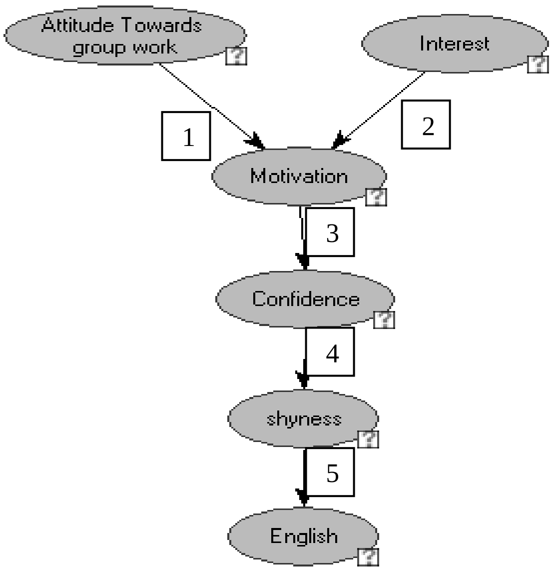
1 ___ Yes _____ No
2 ___ Yes _____ No

(iv)



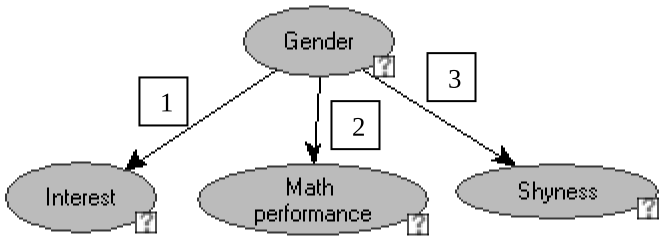
1 ___ Yes _____ No
2 ___ Yes _____ No

(v)



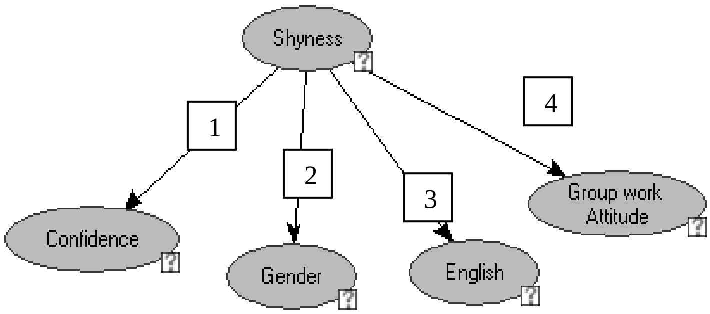
1 ___ Yes _____ No
2 ___ Yes _____ No
3 ___ Yes _____ No
4 ___ Yes _____ No
5 ___ Yes _____ No

(vi)



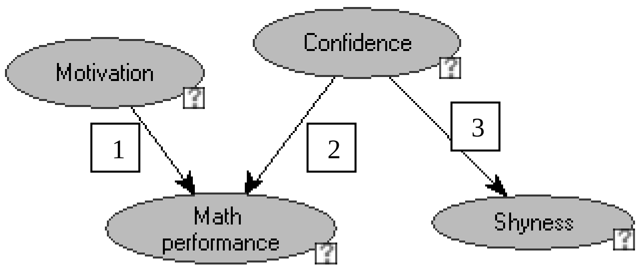
1 ___ Yes _____ No
2 ___ Yes _____ No
3 ___ Yes _____ No

(vii)



1 ___ Yes _____ No
2 ___ Yes _____ No
3 ___ Yes _____ No
4 ___ Yes _____ No

(viii)



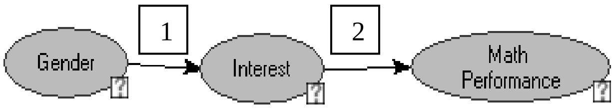
6 ___ Yes _____ No
7 ___ Yes _____ No
8 ___ Yes _____ No

APPENDIX - F
Responses of experts from the survey on network modifications

No	Dependencies	Relative Frequency of positive answers
1	Group work Attitude affects confidence	15%
2	Confidence affects Group work attitude	55%
3	Confidence affects extent of mathematics performance	92%
4	Confidence affects motivation	46%
5	Confidence affects shyness	85%
6	English ability affects extent of mathematics performance	90%
7	English ability affects group work attitude	26%
8	Gender affects extent of shyness	75%
9	Gender affects extent of mathematics performance	85%
10	Gender affects interest for math	85%
11	Interest for mathematics affects confidence	33%
12	Interest affects extent of mathematics performance	100%
13	Mathematics performance affects extent of confidence	45%
15	Mathematics performance affects extent of English performance	0%
16	Mathematics performance affects group work attitude	18%
17	Mathematics performance affects interest for math	58%
18	Achievement motivation affects Attitude towards group work	80%
19	Achievement motivation affects confidence	76%
20	Achievement motivation affects extent of mathematics performance	80%
21	Shyness affects extent of confidence	85%
22	Shyness affects extent of English performance	88%
24	Shyness affects group work attitude	80%

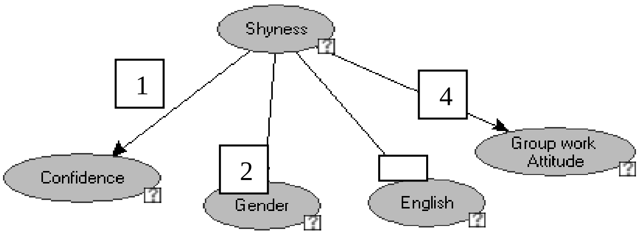
II. A response summarized on the causal diagrams.

(i)



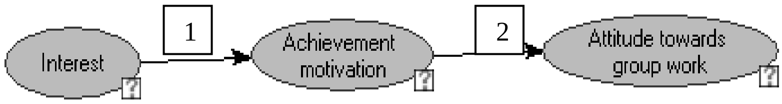
1 85%
2 100%

(ii)



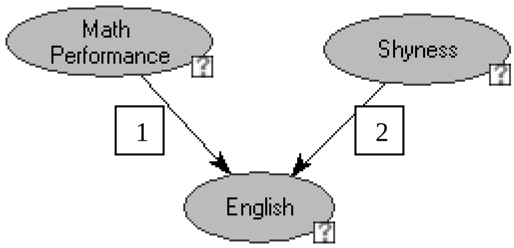
1 85%
2 0%
3 88%
4 80%

(iii)



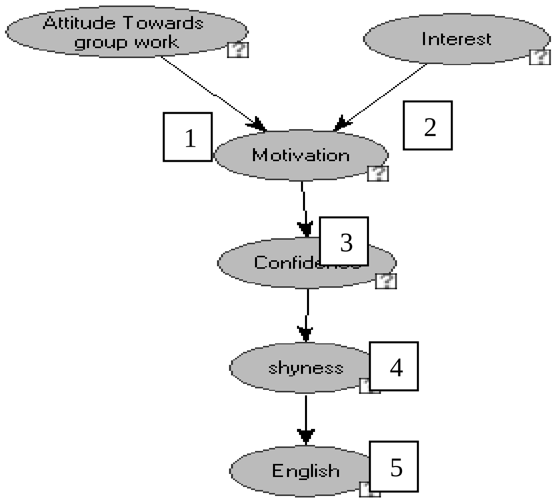
1 0%
2 80%

(iv)



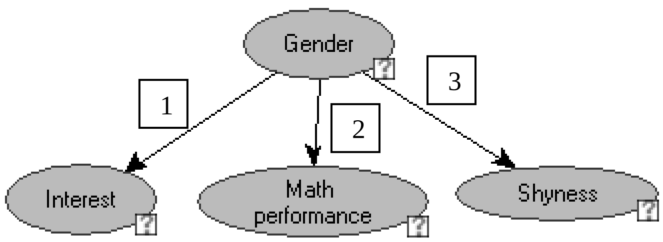
1 0%
2 88%

(v)



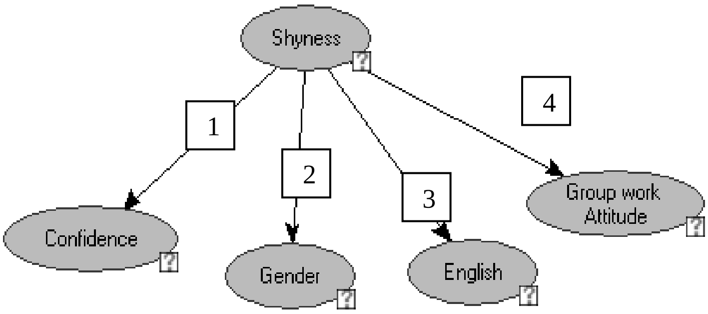
1 10%
2 5%
3 76%
4 85%
5 88%

(vi)



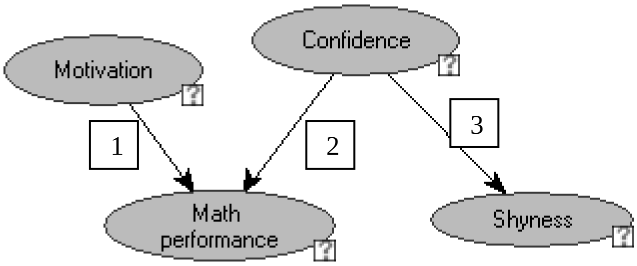
1 85%
2 85%
3 75%

(vii)



1	85%
2	13%
3	88%
4	80%

(viii)



1	80%
2	92%
3	85%

APPENDIX - G
Mathematics exam administrated to students
(Post-group work)

Name: ____ section ________

1. Determine which of the following algebraic expressions are polynomials

	Polynomial	Not a Polynomial	Why
a. $x^4(x^5-\sqrt{5})$	$\square$	$\square$	____________
b. $(x-1)(x^2+1)$	$\square$	$\square$	____________
c. $\frac{6}{x^3}+\frac{5}{x}-15$	$\square$	$\square$	____________
d. $-5x^4+3\sqrt{x}+5$	$\square$	$\square$	____________

2. If f(x) = $6x^6+\sqrt{3}x^2+4$ and g(x) = $-2x^2+3$: find f(x) . g(x)

3. Find the remainder when f(x) is divided by g(x) where
   f(x) = $3x^4-6x^3+2x-1$ g(x) = $x+1$

4. Simplify the following $\left(\frac{8a^{-4}b^3}{27a^2b^{-3}}\right)^{\frac{1}{3}}$

5. Use the sign chart or algebraic method to solve the following
   a) $\frac{x^2+x-2}{x^2-2x-3}\leq 0$

6. Find the lowest terms of the following numbers.
   a) $\frac{65}{13}$ b) $\frac{58}{12}$

7. Solve for x and indicate the domain.
   a) $3-\left(\frac{2x-3}{3}\right)=\frac{5-x}{2}$ b) $\frac{2x-3}{9}-\frac{x+5}{6}=\frac{3-x}{2}+1$

8. Write the following in exponential notations
   a) $\log_6^x=3$ b) $\log_{10}^{100}=2$ c) $\log_y^{150}=3.2$

9. Write the following in Logarithmic Notations
   a) $4^{2x}=256$ b) $2^{\frac{5}{2}}=N$ c) $7^{5x+4}=49$

APPENDIX - H

Post-group work questionnaire

(Translated from Amharic version)

Dear Student,

This questionnaire is prepared to get your feed back on the group work you have been attending for the last eight weeks. Please feel free to indicate your true feelings for each of the questions provided. The researcher would like to take this opportunity to thank you for your participations in the group work.

1. In your opinion the group size was
   - [ ] small
   - [ ] Large
   - [ ] average/well
2. If you were to select your own group, your preference would have been
   - [ ] working with friends
   - [ ] working with high performers
   - [ ] working with low performers
   - [ ] do not really care
3. In your opinion, how do you see the extent of participation of the group members?
   - [ ] all have participated
   - [ ] only few have participated
   - [ ] none of us participated
4. What is your assessment on the extent of efforts to do each question accurately?
   - [ ] every effort is put
   - [ ] not much effort
   - [ ] no effort at all
5. Would you be willing to work with the same group again?
   - [ ] Yes
   - [ ] No
   - [ ] I don't mind
6. In your opinion, what benefits did you gain from the group work?
   - [ ] learnt other ways of solving problems
   - [ ] made friends with others
   - [ ] learnt to help others
   - [ ] improved my abilities to work with others
   - [ ] didn't get any advantage
7. Do you think the group work was successful?
   - [ ] Yes
   - [ ] No
8. In your opinion, what were the unfavourable practices during group work?
   - [ ] members didn't listen to each other
   - [ ] some were loaded with responsibilities
   - [ ] I was bored of regular meeting with the group
   - [ ] I didn't like the group

Appendix I(i)

Algorithm – 1: Considering the First Student as Group Kernel - Pseudo code

(a) First Module (First scan) – beginning of group formation

<u>INPUT</u>: *Outlier-File*
<u>OUTPUT:</u> *Grouped-File* and *Ungrouped-File*
<u>PROCESS:</u>

```
1.    Open the Outlier-File
2.    Open Grouped-File and Ungrouped-File
3.    Create Unclustered-File (temporary file)
4.    initialize GroupNo=1
5.    i = 0   MemberCount = 0
6.    Make the First student (S1), from the Outlier-File, the group initiator.
7.    i = i +1
8.    until end of Outlier-File file
      {
         a.  Read the next student (Si),
         b.  Apply difference measure (d) on Si and the group initiator(S1)
         c.  If d >= Predefined Threshold ,
          {      Put Si in Grouped-File,
                        i = i +1
                 MemberCount = MemberCount + 1;
                 If  MemberCount < 3 then
                         Go to Step (8)
                       Else
                            Append S1 (the group initiator) to the group
                   GroupNo = GroupNo +1;
                           Append remaining Records of Outlier-File to Unclustered-File
                   Make Unclustered-File Outlier-File
                   Go to Step (5)
          }
          If d < Predefined Threshold
          {      Append Si to Unclustered-File
                        i = i +1
                 Go to Step (8)
          }
      }
9.    If MemberCount > 0
          {      Append S1 (the group initiator) to Grouped-File
                 GroupNo = GroupNo +1
          }
      Else
          Append SI (the group initiator) to the Ungrouped-File

10.   If number of record in Unclustered-File > 1
          {      Make Unclustered-File Outlier-File
                 Go to Step (5)
          )
       Else
          {      If number of record in Unclustered-File is = 1
                         Append the record to Ungrouped-File
                 Make Ungrouped-File Outlier-File
                 End
          }
```

Appendix I(ii)

Inclusion of Outlier Students - Pseudo code

(b) Second Module (Second scan) - Outlier Inclusion-I

INPUT: *Grouped-File* and *Outlier-File*

OUTPUT: Updated *Grouped-File and Updated Outlier-File*

PROCESS:

1. Open the *Grouped-File*
2. Open the *Outlier-File*
3. ***i*** = 1
4. While there are groups formed in the *Grouped-File*
    a. For each Group
        If ***MemberCount*** < 4
        {
            i. If End-of-File of *Outlier-File* is Reached
                a. End
            ii. Make the ***i**^th^* record (S_i) in *Outlier-File* group initiator
            iii. ***i***= ***i*** +1
            iv. Apply the difference measure (d) between S_i and **each** student in the Group;
            v. If d >= Predefined Threshold ,
                {
                    Remove S_i from *Outlier-File*
                    Append S_i to the Group
                    MemberCount = ***MemberCount*** + 1
                    If ***MemberCount*** <4
                    {
                        Go to step (**a**)
                    }
                }
        }

Appendix I(iii)

Outlier Exchange - Pseudo code

(c) Third module (third scan) - Outlier Exchange (This works if there are still some students in the Outlier file)

INPUT: *Grouped-File* and *Outlier-File*
OUTPUT: Modified *Grouped-File* and Modified *Outlier-File*
PROCESS:

1 Open the *Grouped-File*
2 Open the *Outlier-File*
3 ***i****=1*
4 While there are groups in the *Grouped-File*
   a. For each Group
      i. If End-of-File of *Outlier-File* is Reached
         1. End
      ii. Exchange i^{th} record from the *Outliers-File* with the first record of the group in the *Grouped-File*
      iii. Make the new record the group seed
      iv. For the remaining Group members
         1. Compute the difference measure (***d***) with the group seed
            If ***d***>= Predefined Threshold ,
            {
               - Append the New Record to the group
               - Add the Exchanged Record from the Group to the Outlier File
               - Go to Step (***a***)
            }
            Else
            {
               i = ***i*** +1
               Go to Step ***(i)***
            }

Appendix I(iv)

Grouping outlier students - Pseudo code

(d) Fourth Module (fourth scan) - make groups with the outlier file (Exchanged students list)

INPUT: *Outlier-File and Grouped-File*

OUTPUT: *Modified Grouped-File* and Modified *Outlier-File*

PROCESS:

1. Open the *Outlier-File* (now it is the exchanged students)
2. Open the *Grouped-File* with the option of appending groups
3. Run **(a) First Module (First scan)**

(e) Fifth Module (Fifth scan) – examining final group candidates

INPUT: *Outlier-File and Grouped-File*

OUTPUT: *Final Grouped-File* and Modified *Outlier-File*

PROCESS:

1. Open the *Outlier-File* for appending
2. Open the *Grouped-File*
3. Open *Final Grouped-File* for Appending
4. Until end of *Grouped-File*
    1) For Each Group in *Grouped-File*
        - If all group members have low values for a specific attribute then (the group can not be included in the final cluster)
            - Remove from grouped and append all members in *Outlier-File*
        - Else If the group average is less than the group threshold (eqn. 5.6) then (the group can not be included in the final group)
            - Remove from grouped and append all members in *Outlier-File*
        - Else
            - Append the Group to *Final Grouped-File*

Note:

- the fourth pass creates group with the outlier file
- The fifth pass selects the final groups and drops those groups which do not meet the criteria specified.

Appendix I(v)

Repetition of previous modules - Pseudo code

(f) Sixth Module (Sixth scan) – Repeat Steps [a] up to [e]

INPUT: *Outlier-File and Final Grouped-File*

OUTPUT: *Modified Final Grouped-File* and Modified *Outlier-File*

PROCESS:

1. Until end of Outlier-File
2. ***i*** = 1
3. ***Outlier[0]=*** *Outlier-File*
4. Until No Record in *Outlier-File*
   - Do
     {
     - a. Steps **(a) First Module– (e) Fifth Module**
     - b. Name Outlier file ***Outlier[i]***
     - c. If ***Outlier[i]*** Different from ***Outlier[i-1]***
       {
       - ***Outlier[0]= Outlier[i]***
       - ***i = 1***
       - ***Go to Step (a)***

       }
     - d. Else
       {
       - ***i*** = ***i*** +1
       - if ***i=40*** Then ***End***
       - Else ***Go to Setp (a)***

       }

     }

Note:

- This is the repletion steps until all students are finally grouped, or there are no more outliers.

Appendix I(vi)

Appending outlier students - Pseudo code

(g) Seventh Module (Seventh scan) – Outlier Inclusion-II

INPUT: *Outlier-File and Final Grouped-File*
OUTPUT: *Modified Final Grouped-File* and Modified *Outlier-File*
PROCESS:

1. Open the final *Grouped-File*
2. Open *Outlier-File*
3. $\boldsymbol{i} = 1$
4. Until end of *Outlier-File*
   i. Take the $\boldsymbol{i}^{th}$ record ($\boldsymbol{OS_i}$) from the *Outlier-File*
   ii. For Each Group
   {
   If ***MemberCount*** < 4
   {
   a. Include the $\boldsymbol{i}^{th}$ record from *Outlier-File to the Group temporarily*
   b. If Group-Average < Group-Threshold (eqn 5.6)
   - *Remove* $\boldsymbol{OS_i}$ from the Group
   - *Go to Step (**ii**)*

   c. Else
   - Leave $\boldsymbol{OS_i}$ in the Group
   - $\boldsymbol{i} = \boldsymbol{i} + 1$
   - if ***MemberCount*** < *4*
     Go to step ***(a)***

   }
   }

(h) Eighth Module (Last scan) – Forcing Outliers

INPUT: *Outlier-File and Final Grouped-File*
OUTPUT: *Modified Final Grouped-File*
PROCESS:

1. Open the final *Grouped-File*
2. Open *Outlier-File*
3. $\boldsymbol{i} = 1$
4. Until end of *Outlier-File*
   a. Take the $\boldsymbol{i}^{th}$ record from the *Outlier-File*
   b. Include *the* $\boldsymbol{i}^{th}$ record to the $\boldsymbol{i}^{th}$ Group
   c. $\boldsymbol{i} = \boldsymbol{i} + 1$
   d. Go to step (***a***)
5. Stop.

Note:
- These two modules append outlier students who could not be assigned to any group.

Appendix I(vii)

Algorithm – 2: Considering the Last Member as Group Kernel - Pseudo code

(a) First Module (First scan) – beginning of group formation

<u>INPUT</u>: *Outlier-File*
<u>OUTPUT:</u> *Grouped-File* and *Ungrouped-File*
<u>PROCESS:</u>

1. Open the *Outlier-File*
2. *Open Grouped-File* and *Ungrouped-File*
3. initialize ***GroupNo***=1
4. ***i*** = 0 ***MemberCount*** = 1
5. Make the First student (***S1***), from the *Outlier-File*, the group initiator of the 1st group.
6. ***i*** = ***i*** +1
7. While number of records in outlier file > 2

```
{
    a.  Read the next student (Si),
    b.  Apply difference measure (d) on Si and the group initiator
    c.  If d >= Predefined Threshold ,
     {        Put Si and S1 in Grouped-File
                    Make Si the group initiator ( S1= Si )
                    i = i +1
              MemberCount = MemberCount + 1;
              If  MemberCount < 4 then
                    Go to Step (a)
                  Else
                GroupNo = GroupNo +1
                i = i +1
                    Make Si the group initiator ( S1= Si )
                Go to Step (a)   }
     If d < Predefined Threshold
     {        Append Si to Unclustered-File
                    i = i +1
              Go to Step (a)   }
```

8. If MemberCount = 1
   Append ***S1*** (the group initiator) to *Ungrouped-File*
9. If number of record in Unclustered-File > 1
   { Make *Unclustered-File Outlier-File*
   Go to Step (***5***) }

Appendix I(viii)

Algorithm – 3: Considering a Low Performer as Group Kernel - Pseudo code

(1) (a) First pass/first scan– Begin group formation

INPUT: *Outlier-File*, **OUTPUT:** *Grouped-File* and *Ungrouped-File*

PROCESS:

1. Open the *Outlier-File*
2. *Open Grouped-File* and *Ungrouped-File*
3. initialize ***GroupNo***=1
4. ***i*** = 0 ***MemberCount*** = 0
5. until ***mathematics result*** < ***satisfactory*** for the i^{th} Student
   i. ***i*** = ***i*** +1
   ii. Read the i^{th} Student from the *Outlier-File*
6. Reorganize the Outlier File to make S_i the first student
7. ***i*** =0
8. Make the i^{th} Student the group initiator
9. ***i*** = ***i*** +1
10. While number of records in outlier file > 1
    a. Read the next student (S_i),
    b. Apply difference measure (***d***) on S_i and the group initiator (**equation 5.3**)
    c. If ***d*** >= Predefined Threshold ,
    { Put S_i in *Grouped-File*,
    i = ***i*** +1
    MemberCount = ***MemberCount*** + 1;
    If ***MemberCount*** < 3 then
    Go to Step (***a***)
    Else
    Append the group initiator to the group
    GroupNo = ***GroupNo*** +1;
    Append remaining Records of *Outlier-File* to *Unclustered-File*
    Make *Unclustered-File Outlier-File*
    Go to Step (***4***)
    }
    If ***d*** < Predefined Threshold
    { Append ***Si*** to *Unclustered-File*
    i = ***i*** +1
    Go to Step (***a***)
    }
11. If ***MemberCount*** > 0
    { Append the group initiator to *Grouped-File*
    GroupNo = ***GroupNo*** +1
    }
    Else
    Append ***SI*** (the group initiator) to the *Ungrouped-File*
12. If number of record in Unclustered-File > 1
    { Make *Unclustered-File Outlier-File*
    Go to Step (***4***)
    }

    Else
    { If number of record in *Unclustered-File* is = 1
    Append the record to *Ungrouped-File*
    Make *Ungrouped-File Outlier-File*
    End}

Appendix I(ix)

Incremental version for Group composition - Pseudo Code

The Incremental Algorithm

INPUT: *New Student, previous-grouped-file*

OUTPUT: *updated-grouped-file*

PROCESS:

1. ***MemberCount*** = 0
2. ***GroupLabel*** =1
3. For Each New Student (***S***) {
   - (i) Collect required information from ***S***
   - (ii) Measure the Attributes
   - (iii) Predict Mathematics Performance of the Student
   - (iv) If previous-*grouped-file* is empty {
     - a. Append ***S*** to the *Grouped-File,*
     - b. Make the Student (***S***) the group Seed.
     - c. Assign ***GroupLabel*** for the Group
     - d. Output updated-grouped-file}
   - (v) Else {
     - i. For Each Existing Group in the previous-grouped file
     - ii. Count the number of Members
     - iii. For All Groups where ***MemberCount*** < 4
       - i. Apply difference measure (***d***) on ***S*** and all group initiators
     - iv. If there exists a Group where ***d***>= Predefined Threshold {
       1. For all Groups where ***d* > =** Predefined Threshold {
          - a. For all Groups where ***MemberCount*** = 3 {
            - i. Add ***S*** to the group **IF**
              - *Group-Average* > *Group-Threshold* ***AND***
              - No attribute in the group having same low value for all members ***AND***
              - ***d*** is the larges from all groups in this category
            - ii. end } }
       2. For all Groups where ***MemberCount*** < 3

          {

          - a. Add ***S*** to the group where ***d*** is the largest from all groups in this category.
          - b. Make ***S*** the Group Seed in the Group

          } }
     - v. Else {
       1. Make a new Group where ***S*** is the group initiator
       2. Assign a new Group label for the Group
       3. Output updated-grouped-file

       } } }

Appendix J

Some pictures of students[41] attending group work

[41] Permission was obtained from students for their picture to appear in this thesis.

Ich erkläre hiermit an Eides statt, dass ich diese Arbeit selbst verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe.

Hamburg, im Oktober 2005

Rahel Bekele