CRANFIELD INSTITUTE OF TECHNOLOGY

CENTRE FOR TRANSPORT STUDIES

PhD THESIS

Academic Year 1989-90

Lawrence Harrell

The Effect of Intangible Product Attributes on Rail Passenger Demand with Special Reference to Ride Quality

Supervisor: Ian Black

March 1990

To Mum and Dad

Abstract

This study set out to investigate the value consumers place on less tangible product attributes. Although some work has been done in the past, to establish the relative importance of intangible attributes; very few studies have attempted to produce a financial value for a change. The research was conducted in a rural railway environment and so the product considered was a train journey. The main intangible attribute chosen for the study was ride quality.

Rural railway services make significant losses and as a result have been threatened with closure. Reducing track maintenance (and thus ride quality) on these routes offers considerable scope for cost reduction. But, very little was known about the response of demand to changes in ride quality. Any results obtained could, therefore, make a contribution to maintaining railway services in areas of limited public transport.

Although this study concentrated on the railway ride problem, it is believed that the method developed during this research would be applicable, with some modification, to other topics.

Acknowledgements

I wish to acknowledge the support of the following individuals and organisations, without which this research would never have been completed.

My supervisor Ian Black, for guiding me through a sometimes frustrating process and for his excellent proof-reading.

All the staff and students at C.T.S, who have been involved in the research, especially Dr. John Towriss.

I would like to thank all those in British Rail who assisted me in the project. Dr. David Round (of the Railway Technical Centre, Derby) made considerable efforts throughout the research. Peter Dunkerly (also R.T.C, Derby) for his help in setting up the research. Peter Ryan (ex British Rail Provincial, Swindon) for his help in setting up the interviews on the Bristol-Weymouth line. All the guards who worked on the Bristol-Weymouth line during 1989, for making the data collection so enjoyable and trouble free.

I would like to thank my parents for supporting me all through my extended education. My grandparents and great aunt for their financial assistance. Also my sisters Katrina and Vanessa.

I would like to thank Amanda Rolfe for enduring the troubles of the first year. To all the Waterloo Boys: Kevin "Hippy", Dave "Thug", Matt "Cyclist", Andy "Farmer", Martin "Tart", Daffyd and Soapy for keeping me sane during the last half of the research. Thanks especially to Kevin for his psychological and multivariate statistical advice, also to Daffyd for his help with proofing.

Finally, my thanks to Ami(ga) for the long hours of work with few complaints.

Contents

Chapter One Introduction

1. GENERAL:

This research investigates peoples' views on the design of rural railways. The method developed in the work should be applicable in many other environments.

In this research rural railways are defined as: lines that are not dominated by limited stop inter-city services or commuter flows. Such rural lines are known in British Rail as, "Other Provincial Services". The research considers intangible product attributes, defined as: attributes for which there is no ubiquitous form of measurement. A main objective is to establish the financial value of a change in the level of such attributes, from now on this is referred to as the value of the attribute.

2. THE RURAL PUBLIC TRANSPORT PROBLEM:

A number of authors (for example: Hillman and Whalley (1980), Moseley (1979), Nash (1982) and St John-Thomas (1963)) have described the problems of rural transport.

Everyone needs accessibility, which Moseley (1979) defines as, "Peoples' ability to reach the things that are important to them". The need for accessibility in rural areas is heightened by, "The inadequacy of employment opportunities, selective depopulation and repopulation, the isolation and loneliness of certain vulnerable groups and the dis-proportionately high cost of providing services" (Moseley (1979)).

Providing rural public transport, to meet this need, has always been difficult. The demand for rural transport is low and the population widely spread, with people needing to visit the larger concentrated centres. This reduces the viability of public transport operation, as conventional forms are too inflexible to effectively tap such dispersed demand. Rural railways, with their fixed single track are particularly bad in this respect.

Overall, the level of transport available in rural areas has never been as good as it is now, because the private car has had a liberating effect on most of the community. But for those without cars, mobility is restricted and in some cases no public transport is available at all. Moseley (1979) states that, "Accessibility is very unequally distributed in rural areas. Not only do places or villages differ considerably in this respect, but so, too, do individual people".

According to the Government Statistical Service (1989) 36% of households do not have the use of a car - though this figure is lower in rural areas (for example, Transport and Road Research Laboratory (1980)). The people in the community who tend to be deprived, in this sense, are concentrated among: the poor, the old, children and housewives - who are left without the family car when the husband goes to work. These groups are, in many ways, those who can least afford to be isolated and this must detract from their enjoyment of rural life.

These problems have been exacerbated in recent years by a number of trends (for example, Nash (1982)):

a). Continued growth in car ownership has reduced the already limited custom for public transport. A less direct effect of increased car ownership has been a reduction in the profits made on urban and inter-urban routes. Traditionally this surplus has been used to cross-subsidise rural services - this is no longer possible.

b). A rise in the real cost of rural service provision has meant that fewer rural settlements now have facilities such as post offices and shops; these services being concentrated in larger villages and towns.

c). The trend of rural de-population has continued in many of the least accessible areas, further reducing demand for public transport.

d). An ageing population and greater traffic hazards, mean that walking and cycling are less practical than they were in the past.

Due to these problems, it has long been recognised that public transport has to be subsidised to some degree. However experience has shown, that without careful management, the cost of provision can escalate to a point where the service is cut back or withdrawn completely (for example, Hamilton and Potter (1985)).

3. RURAL RAILWAYS:

3.1. Basic Problem:

Rural railways have always been associated with financial difficulties. During their heyday, when even the remotest areas were served, it was doubtful whether much of the system made a profit (for example, Keen (1978) and Kilvington (1983)). Despite this, little was done to change the pattern of services until the 1960's.

The basic problem with rural railways is the large infrastructure overhead, typically accounting for 30-40% of costs (for example, Nash (1982)). With rural services there are few passengers to contribute towards these fixed costs and so the service makes a loss. The low potential demand is reduced further by the infrequent service and thus inconvenience, associated with it.

3.2. Background to the Problem:

Many authors have described the haphazard development of rural railways (for example: Freeman-Allan (1985), Gammell (1983), Hedges and Whitehouse (1980), St John-Thomas (1977), St John-Thomas and Whitehouse (1986) and Nock (1957)).

Most rural railways were constructed during the late 19th century and formed the first large scale public transport system in these areas. The lack of a coherent planning framework, meant that a significant number of these lines were not built primarily to serve existing settlements; but rather to stop a competitor gaining (sole) access to major centres, for civic pride, or as speculative ventures. Further, the whims of aristocratic landowners could determine the alignment of routes. The results of these policies have important consequences today: some settlements have two or more unconnected stations on opposite sides of the town, or stations several miles from the settlement they are supposed to serve.

Rural railways began to lose traffic in the 1930's with the development of bus services. On Nationalisation in 1948, British Rail took control of the large number of rural routes. The network and its operation had changed little since construction. By the late 1950's the private car began to have an impact and the network became very unprofitable. Although some attempts were made to rationalise the system, it was not until Dr. Beeching's report in 1963 that any real progress was made. The report stated that, in 1961, one half of the network carried only 4-5% of total traffic and that the railway had become almost insignificant in rural areas (British Railways Board (1963)).

When the report's recommendations were implemented, the rail network shrank from 20,237 miles in 1963, to 15,242 miles by 1968. But, before the report had its full effect the 1968 Transport Act began to question its implementation. The social consequences of rail closures were now also to be considered. As a result, some lines and services scheduled for closure under the Report have survived, notably those in the Scottish Highlands.

From the Beeching era until the early 1980's, the lines that remained were gradually run-down to reduce costs. Stations were unmanned, capacity was reduced, investment and maintenance were cut. While these policies reduced the cost of operations they tended to make the services less attractive, reducing patronage still further.

By the early 1980's it was clear that major investment would be needed to prevent the widespread withdrawal of rural services. By 1982 British Rail's rural services were loosing 17.9 pence per passenger mile and only 1% of rural services were covering their direct operating costs (Kilvington (1983)).

During this period a series of studies were conducted that suggested solutions to the problem (Association of County Councils et al (1983), Central Transport Consultative Committee (September 1979), Policy Studies Institute (1981) and the Serpell Committee (1983)). The Serpell Committee was the most radical and proposed a series of cuts to the network. However, these recommendations produced such an outcry, that the Government (publicly) abandoned the idea of any significant cuts in the network. The Government thus authorised sufficient investment to remove any immediate threat.

In 1982 the whole structure of British Rail's management changed, to make it more business orientated (Reid (1982), Nash (1985)). Since this change, the responsibility for rural railways has been clearly assigned to British Rail's Provincial Sector.

Rural railways are currently supported by the P.S.O. (Public Service Obligation) grant, in return for which British Rail are expected to run a service comparable to that of 1975. Local authorities can, and have, topped up the P.S.O. grant to enable British Rail to run additional services on a marginal cost basis.

During the 1980's the Government have consistently reduced the P.S.O. grant to Provincial. Provincial have had considerable success in reducing their losses to the new grant levels.

This financial improvement has taken place both through cost reduction and taking better advantage of market opportunities. This is described by, for example, Railway Gazette International (May 1985) and Abbott (March 1986).

The increase in marketing activity is illustrated by the large amount of promotional literature produced by the Sector, for example: (Provincial (1982, 1985, 1986, 1987, 1987)).

Cost reduction has been achieved through the introduction of new more efficient rolling stock - Sprinters and Pacers (British Railways Board (1985 and 1986), Ford (June 1986), Haresnape (1985 and 1986), Perren (June 1986), Rail Engineering International (1985), Rail Power (October 1985 and July 1986) and Railway Gazette International (November 1981, July 1982, January 1983 and July 1984). These new trains have not been without their problems, being more cramped than those they replaced and having greatly reduced luggage space. The Pacers in particular have proved unreliable (David (January 1988) and Beresford et al (1986)).

New signalling practices have greatly reduced the amount of manpower required (for example: Apperson (February 1987) and Railway Gazette International (February 1981 and November 1982)).

The current financial position is depicted in British Rail's Annual Report. Provincial's gross income in 1988/9 was £274.1 million. Losses in 1988/9 were £465.9 million, reduced from £661 million in 1984/5 (1989 prices). Total operating expenses per mile have been cut, from £10.31 in 1984/5 (1989 prices) to £8.66 in 1988/9 (British Railways Board (1989)).

In 1986 Provincial accounted for 58% of British Rail's stations, 34% of train miles and 53% of route miles - illustrating the relatively low intensity of the use of its track (British Rail Provincial (April 1986)). It should be noted that these figures include commuter services outside London, which are also part of Provincial - if urban services are removed the figures are lower.

3.3. Bus/Coach Replacement of Services:

The problems of railways in rural areas, mean that their replacement by buses or coaches is often suggested. Public road transport is more flexible, with a much smaller overhead and so should be more viable in rural areas. For example, Keen (1978) cites the advantages of bus substitution to be, "Smaller vehicle, freedom from scheduling limitations of single lines and junctions, the ability to penetrate the market square and detour to the village".

Bus/coach replacement of rail services was attempted on a wide scale after the Beeching cuts, but was not always successful. So few passengers transferred to the substitute services, that many were abandoned after only a few years with no closure enquiry (Hamilton and Potter (1985), Hillman and Whalley (1980)). Even more recent cases of substitution, have proved unsuccessful. For example, the bus service that replaced the Bridport to Maiden Newton line (closed in 1975) was within a few months of closure carrying, "Only one quarter of the ex-rail passengers" (Hamilton and Potter (1985)).

A number of reasons have been proposed for the past failure of bus/coach substitution. For example, the Railway Development Society (1977) lists: limited luggage space - especially for prams and cycles, a poorer ride, longer journey times, inadequate connections with remaining rail services, no toilets, cramped conditions, inability to read/work and unreliability during bad weather and summer congestion. The flexibility of road services has tempted operators into serving additional places along the route, further increasing journey times.

Some still advocate bus/coach substitution, arguing that the problems of the past were only a result of poor implementation. It is suggested that bus/coach replacement would be more effective, if the services were operated by British Rail, allowing integration with the rail network. Keen (1978) argues that, "The failures alleged and in considerable measure real - of past closures have, in my view, very little to do with the change of mode".

Examples of alternative bus/coach services at an eighth of the cost of current rail services are still quoted. But, many studies assume the same number of passengers will use the substitute service, are based on generalised railway costings and compare a poorly run railway with an efficient bus/coach operation (for example, Macbriar (July 1983)).

The Serpell Report suggested that, if certain conditions were met, a widespread bus/coach replacement of rural rail services would be desirable (Serpell Committee (1983)). One of these conditions was that British Rail should have some control over replacement services. British railway companies have operated road services in the past, but these have been sold-off - this could happen again in the future.

Serpell argued that replacement services should be more secure and subsidised by Central, rather than Local, Government. But, replacement road services can still be withdrawn without closure proceedings (for example, Hamilton and Potter (1985)) and so do not currently offer a secure alternative service. The latest Monopolies and Mergers Report was not convinced of the benefits of replacement services (Monopolies and Mergers Board (1989).

3.4. Justification for Rural Rail Services:

Apart from the need to ensure mobility for the disadvantaged and the problems with bus/coach substitution, there are a number of reasons why rural rail services are maintained (for example, Railway Development Society (1977 and 1984)).

The cost of reinstating a railway means that closure is nearly always irreversible. A bus/coach service can be reinstated quickly and at a minimal cost. There have been occasions where, with hindsight, closure should not have occurred. New traffic flows could result from the development of new settlements, discoveries of natural resources, or re-routing of inter-urban services.

Some rural railways have been shown to produce significant contributory revenue for primary routes (Polytechnic of Central London (February 1976)). The closure of rural services could cause previously remunerative mainlines to become unprofitable.

Rural railways sometimes provide journey times that cannot be matched by an alternative bus/coach service, without considerable investment in new roads. Examples of this can be seen at Gunnislake in Cornwall (St John-Thomas (date unknown)) and the Cambrian Coast in Wales (Department of Transport (1969)).

Rural lines in some areas make an important contribution to the local economy, by bringing in tourists during the summer. A number of these lines are used as a leisure attraction, as well as a form of transport (especially in the Scottish highlands). An alternative bus/coach service would not offer this attraction. Such lines often relieve congested narrow roads during the summer and help to contain rural depopulation.

Many lines have survived because they serve government, or military installations. Examples could be, ports and nuclear power stations. It is also argued that some lines have survived, because they serve marginal constituencies.

Few of these reasons are overwhelming and so it is important that losses are contained, if services are to be maintained.

3.5. The Outlook and Need for a Valuation of Ride:

Further reductions in costs are therefore required to secure the future of the rural rail network. The most effective way of achieving this, is to reduce expenditure on the fixed overhead. A major component of this is the cost of track maintenance, any change in which will affect the quality of ride offered to customers (and consequently the number of people who use the service).

The new trains, recently introduced on rural lines, have much better suspensions. This means that rural rail passengers now experience a significantly better ride. It may now be possible to reduce track maintenance on these lines, while still providing the user with a better level of ride than with the old rolling stock. Even if track quality was reduced by 25% on Provincial routes, the ride would still be better than on InterCity routes - with the latters' higher speeds (Frederick (October 1987)).

British Rail would like to optimise the relationship between ride quality and revenue. A decision therefore has to be made on how far to reduce track maintenance, to maximise the profitability of the service.

British Rail (Derby) have developed models that relate track maintenance expenditure, to the ride quality felt by passengers (Frederick (October 1987)). The less obvious results of changes in track maintenance, have also been researched by British Rail. These effects include changes in: the costs of traction energy, the number of derailments, the amount of suspension maintenance and the damage to track components and ballast from changed dynamic loads (Frederick (October 1987) and Round (November 1987)).

British Rail are therefore able to state what the engineering costs of any change in track maintenance are and what the resulting change in ride quality is. They believe that a 30-50% reduction in maintenance costs could be possible, saving £20 million (Round (November 1987)).

What British Rail are not currently able to determine, is the effect of changes in track maintenance (and thus ride quality) on the demand for services. Clearly the rougher ride is made, the more passengers and thus revenue will be lost. British Rail need to know the number of passengers that would be lost from any change, so that the size of this revenue loss can be reliably estimated - this is therefore the main aim of the research.

3.6. The Problems of Valuing Ride:

There are a number of factors that make it difficult to establish a value for a change in ride quality. The main problem is one of scaling. With most attributes that have been successfully valued in the past, respondents are fully conversant with the system of measurement. For example, journey time is measured in hours and fares in pounds. If a researcher asks for a response to a £5 increase in fares, the respondent knows almost immediately how that will affect him.

British Rail measure ride quality using I.S.O. (International Standards Organisation) weighted m/s2 R.M.S. (root mean square). Presenting a change in these units for a respondent to value, is not likely to produce satisfactory results. People do not deal with these measurements on a day-to-day basis and so cannot be expected to comprehend the scales used by engineers. This difficulty of comprehension is exacerbated, as there is no undisputed measure of engineering ride (see chapter two).

If respondents cannot be presented with engineering measures of ride, some other comprehensible form of scaling has to be developed. The only way to achieve this is to produce scales that are relevant to peoples' experience. Such scales might be based on sensations associated with various ride levels, for example: travel sickness or the inability to write.

The problem with alternative sensation scales is that they are open to wide interpretation and so are not directly related to engineering measures. This means that it can be difficult to establish, the change in ride that respondents have valued. One way around this would be to develop a scale based on the ride in a number of different trains previously used by the respondent. The ability of respondents to comprehend such alternative scales is clearly dependent on their level of experience. Care must therefore be taken to ensure that valuation samples have sufficient experience.

Producing a value for a change in ride quality is further complicated, as peoples' perceptions of ride may be interfered with by other attributes - particularly noise. These problems are discussed in detail later in the thesis.

For these reasons, little is known about the effects of intangible product attributes (like ride) on purchase behaviour - this is especially the case in the transport industry. All that is available to transport undertakings are a series of acceptable vibration levels - some of which were produced using less than scientific procedures (see chapter two).

Only one study has attempted to produce a financial value of ride. This was carried out for British Rail's InterCity Sector (M.V.A. (May 1986)). The M.V.A. study did find a value of ride; but the descriptions offered to passengers were so vague, that it is difficult to establish the levels of engineering ride that were valued. These figures were also not designed for rural routes.

4. OBJECTIVES:

The major objective of this research, is to establish a monetary value for a change in ride quality. Other objectives were included: if they would help validate the method, provide insights into the choice process, or would be useful to British Rail.

a). To develop a method that would enable the value of ride (and other intangible attributes) to be estimated, implicit in this are the following sub-objectives.

b). To establish the ways in which ride quality affects passengers and the language they use to describe this.

c). To look at the importance of ride quality relative to other intangible attributes.

d). To investigate the factors that may contaminate passengers' ride assessments and valuations, including the effects of: hyper-sensitivity, recall error, the ability to isolate ride from other intangible attributes, the decision environment, assessment consistency and the sensitivity of perceptions.

e). To establish whether the value of ride is linear.

f). To consider the factors that influence the value of ride and to produce a model to estimate the effects of a ride change, for a number of rail services with different characteristics.

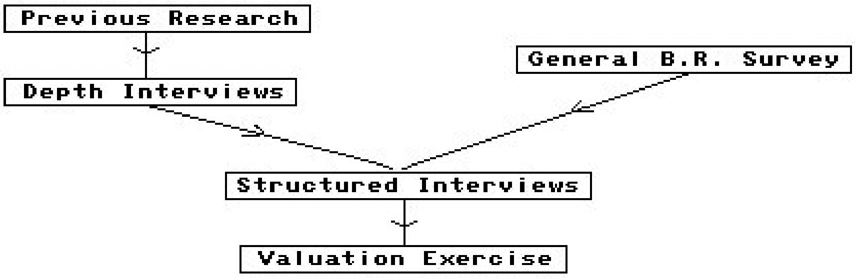
5. GENERAL APPROACH:

The objectives are achieved in a multiphase study.

To successfully establish a value of ride, it is important that more knowledge is obtained about the assessment and valuation of intangible attributes. The first part of the research (chapter two) concentrates on an investigation of the previous work in these areas and draws any relevant information from this, for the development of the research programme.

The next stage is to establish all the possible techniques (chapter three) for estimating a value of ride. Each of these is considered in turn and either eliminated (if it is clearly impractical) or developed. At this stage it is realised, that some form of general investigation into peoples' responses to ride quality is required.

FIGURE 1.1. THE MULTI-STAGE APPROACH.



Once this is completed, the results of a series of depth interviews are analysed (chapter four). This attempts to produce an unambiguous form of language and generally investigate the way in which ride quality affects passengers.

The methodology is then developed in detail (chapter five). All the options that have reached this stage, involve one or two unknowns that have to be investigated, before the techniques can be implemented. These unknowns might be, for example: whether the value of ride is linear, or how well people can remember past ride levels. The areas where further information is required are established and ways of obtaining this considered.

This is followed by the detailed analysis of ride ratings produced in a general quality of service study, conducted for British Rail Provincial, in which the author was involved (chapter six). The relationships between ride and other intangible attributes are considered using this large data set.

Next a series of detailed structured interviews are considered, in which passengers assessed levels of engineering ride (chapter seven). The relationship between ride and other intangible attributes is also studied at this point. Many significant findings are produced. The implications for the design of the ride valuation exercise are discussed. This enables three valuation approaches to be isolated for implementation in the final stage.

Finally the three valuation approaches are developed and implemented, each successfully producing ride values (chapter eight). Results are reported, compared and their significance discussed. These values are considered in the light of the only previous research conducted in this area (M.V.A. (May 1986)).

Finally the research is concluded, the significance of the results is considered and their practical value assessed (chapter nine).

Chapter Two Problem Background

1. INTRODUCTION:

As stated in chapter one, little work has been done to establish the value of intangible attributes: this is especially the case with ride quality. Nevertheless, research has been done in related areas. For example, previous work has been carried out to examine people's response to various levels and axes of movement. Work has also been completed to establish the importance (and sometimes values) of other intangible attributes.

Studying the results of previous research, is likely to indicate the most effective way of achieving the objectives of this thesis. The application of this knowledge, is therefore likely to ensure an efficient development of an approach. The findings of previous research can also be compared to the results produced in the thesis. This may provide insights into the choice process and could give some indication of the validity of the results.

The previous work covered in this review is grouped into the following headings: general ride/vibration research, railway ride/vibration research, railway research into intangible attributes and previous attempts at ride valuation.

2. RIDE QUALITY - ENGINEERING BACKGROUND:

Before going any further, the term, "Ride quality" should be more clearly defined (see: International Standards Organisation (1985), British Standards (1987), Hewgill (September 1974)).

Ride generally refers to the movements felt by a passenger, while travelling in a vehicle. Ride quality is a composite of: vertical, lateral and longitudinal linear accelerations, plus the rotational movements of pitch, yaw and roll. The rate of change of any of these accelerations (jerk) is also important.

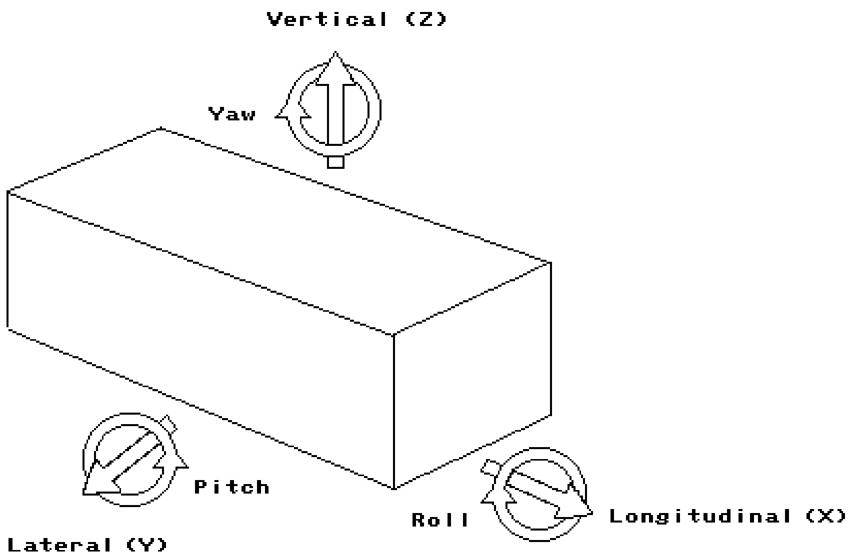
Longitudinal (X) accelerations are usually the result of the vehicle starting and stopping; as rural trains do not generally (de)accelerate rapidly, this movement is of limited significance to the Provincial rail passenger.

Lateral (Y) acceleration is mainly the result of cornering. This is important, as trains can exert strong lateral forces while they are guided round corners by the track.

Vertical (Z) accelerations are caused by bumps and dips in the track. This can also be significant on rail, especially at junctions.

Pitch refers to rotational movement, caused by the up and down motion of the ends of the vehicle: this is usually caused by sudden changes in a vehicles speed. This is not one of the most significant movements on a train.

FIGURE 2.1. COMPONENTS OF RIDE QUALITY.



Roll is a rotational movement, caused by the up and down motion of the sides of the vehicle. Roll can be produced while curving: by the cant on bends and the, lateral "Give" in the coach suspension. This movement can have a significant effect in rail travel.

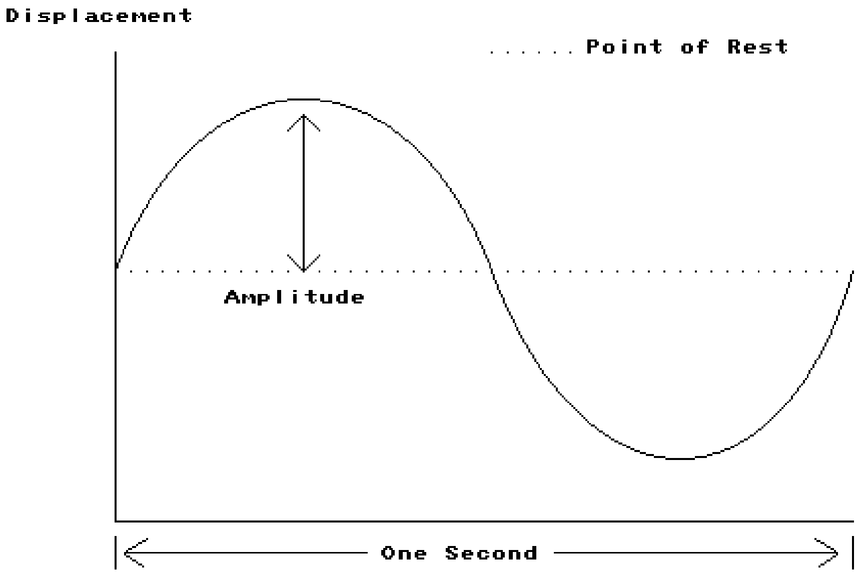
Yaw is a rotational movement, brought about by the side to side motion of the ends of the vehicle. This is of limited significance to the rail passenger.

Acceleration is usually measured in units of metres per second squared (m/s2) or g (9.8 m/s2). Jerk is measured in metres per second cubed (m/s3), with pitch, yaw and roll in units of radians per second. In this paper these measures of ride will be referred to as engineering measures, in order to avoid confusion with the ride ratings developed later. The position of a passenger relative to the axes of rotation, often means that a rotational movement has the same effect as a combination of linear movements. Rotational movements can therefore be partially accounted for, by linear measurements (International Standards Organisation (1985)).

The motion of a vehicle can be considered a combination of different sinusoidal vibrations. Such an analysis assumes that each acceleration is repeated infinitely, so that a series of continuous waves can be used to represent it.

These sinusoidal vibrations are described in terms of their frequency and amplitude (see figure 2.2). The frequency of a vibration, is dependent on the amount of time taken for a vibrating body to complete its cycle and return to its rest position. The faster the vibration, the more cycles will be completed in a second. The number of cycles per second is recorded in Hertz (Hz), which is the standard measure of frequency. Figure 2.2. represents a sinusoidal vibration of 1 Hz. The second important measurement is that of amplitude. This is the maximum amount by which a vibrating body is displaced from its point of rest, at a given frequency.

FIGURE 2.2. VIBRATION MEASUREMENT.



The various frequencies of vibration can be isolated and measured. It has been found that humans are more sensitive to certain frequencies. The British Standard (1987) states that, "The manner in which vibration affects health, activities, comfort and motion sickness is dependent on the vibration frequency". The vibration components at each frequency are thus usually weighted differently. These weights generally follow, either the International (1985) or British Standard (1987) on human exposure to vibration.

Vibrations are usually considered in two or three wavebands. Below 0.5 Hz. they are felt as movement by the passenger, 0.5 Hz. to 25 Hz. are felt as vibrations, while 25 Hz. to 20 KHz. produce sound. This connection between ride and sound, makes it difficult for people to completely isolate the two. This is one of the major difficulties when attempting to produce valid ride ratings.

To be able to analyse vibration over a period, it is necessary to find some way of averaging it. As the movements will be both negative and positive, an arithmetic average would result in a mean of zero. The absolute values of the accelerations could be used, but the most common measure is that of the root mean square (R.M.S). The R.M.S. measure gives increased emphasis to large accelerations.

Although it is clear that the overall ride quality must be some combination of the above vibrations, there is no undisputed way of doing this. The vibrations of each axis therefore has to be reported separately.

3. GENERAL RIDE/VIBRATION RESEARCH:

3.1. General Research into Ride and Comfort:

Early research (for example, Aspinall (1960)) identified the link between the amplitude/frequency of vibrations and passenger comfort. Volunteers were generally driven around a circuit, in cars fitted with accelerometers and gyros - measuring linear and rotational movements respectively. Passengers were then asked for their opinions, which could be compared to engineering measures of ride. As such experiments developed, certain parts of the frequency spectrum were identified as causing more discomfort and levels of equal discomfort were established.

Much of the early research used members of staff to measure perceptions (for example: Aspinall (1960), Aspinall and Oliver (1964), Cooper et al (1978), Cooper et al (1978), Cooper et al (1980)) and they often knew what the experiment was about - resulting in hyper-sensitivity. Either of these factors could give rise to unrepresentatively harsh comments (Cooper et al (1978), Cooper et al (1978), Cooper et al (1980)).

Generally perceptions of ride comfort have been measured using a descriptive scale, for example: comfortable, acceptable, uncomfortable, very uncomfortable (Cooper et al (1978)). Although high correlations have been found between these subjective and engineering measures, "There is a considerable variation in the ratings of riding comfort within each acceleration band" (Cooper et al (1978)).

Some researchers have attempted to go further, producing universal ride quality models that can relate engineering measures to perceptions. Leatherwood et al (1979) shows how difficult this is, as a number of factors interrelate. It was found that the frequency components of ride do not always add - there appears to be no set rules as to how vibrations of various frequencies combine. Noise has been found to interfere with people's perceptions of ride. Under the limited test conditions, Leatherwood et al (1979) found that sound had an additive effect with movement vibration.

A further indication of the interrelationships with ride comes from Stewart (1979). In this study the sensory input to passengers was altered. It was found that those wearing blindfolds, or earplugs gave the best correlations between perceived and engineering ride. Those given a written task found all levels of ride worse, than those who were not. Generally respondents carrying out tasks produced lower correlations. But, Stewart's tests were done using psychology students and not the public.

A bus study (where accelerations are greater than on a train) found that passengers were able to anticipate the accelerations from the road ahead and made stereotyped movements to minimise the effects (Leyland Vehicles et al (1978)). On the modern Provincial trains, passengers cannot see the line ahead and are therefore less able to compensate for movements.

Leyland found that longitudinal accelerations caused significant problems on buses and that the point where passengers were reacting to accelerations was the point where subjective comments changed to uncomfortable. Stabilising poses were adopted by seated, as well as standing passengers. It was found that elderly passengers were often unable to exert enough force on stanchions to remain upright, other passengers remained stable unless the movements were unexpected.

3.2. The International and British Standards:

The British (1987) and International Standards (1985) summarise previous vibration research. They list the effects of vibration, which include: changes in levels of arousal, motivation and fatigue, the acquisition of information via the senses, information processing, interference with activities, discomfort, motion sickness and deterioration of health.

Vision is the sense most affected by vibration. Although suspected, it is difficult to show that vibration affects the cognitive process. Passengers' movements can be made difficult and less precise in the presence of vibration. Low frequency vibration (below 0.5 Hz. - usually in the vertical (Z) axis) can result in motion sickness: characterised by pallor, vomiting and nausea. However, in some cases the effects of vibration can be pleasurable, for example in a fairground.

Factors known to influence human susceptibility to vibration effects, are: age, sex, size, experience, expectation, arousal, motivation, financial involvement, body posture, activities, vibration magnitude, vibration frequency, vibration axis, light, smell and drugs (British Standards (1987), Koffman (October 1968)). Women are more prone to motion sickness than men and young children are particularly sensitive - though immune under 18 months.

British Rail research suggests that passenger comfort (at secondary railway speeds) is best related to the vector sum of International Standards Organisation (I.S.O) weighted vehicle accelerations (Frederick (October 1987)). British Rail thus recommend the use of the I.S.O. proposals on measurement. The I.S.O. state that there is insufficient evidence to produce firm recommendations about measurement - though some universal guidelines are presented.

3.3. International Standard - Measurement Guidelines:

Variation in ratings, both within and between people, mean that the I.S.O. guidelines are based on the average response to vibration. Tables of sensitivity are available for linear accelerations in each of the three axes, though there is little information available on the effects of roll.

The International Standard specifies limits, for a number of frequencies between 1 and 80 Hz. The most sensitive vibration frequencies for humans are in: the X and Y axes below 2 Hz. and in the Z axis between 4-8 Hz. Tolerance has been found to fall with increased exposure, though it is suspected there is some degree of recovery between bouts of vibration.

Several levels of exposure have been established. A reduced comfort boundary (related to reading, writing and eating) has been derived from transport studies and is assumed to lie at about 1/3 of of the fatigue-decreased proficiency level.

According to the International Standard, vibration should be measured at the interface between the body and the supporting structure. The primary quantity used in the measurement of accelerations is m/s2. The magnitude of a vibration should be expressed in R.M.S. If vibrations are very peaky, this standard may not be appropriate. Angular accelerations should be measured in radians per second.

There is no evidence about the interaction of vibrations at various frequencies. So where vibrations of different frequencies occur simultaneously, vibrations should be recorded separately, with the R.M.S. acceleration for each considered relative to the guidelines. If the vibrations are concentrated in a 1/3 octave band (an octave band is double the frequency - for example: 2, 4, 8,... Hz) or less, the R.M.S. of the whole band is taken.

One overall figure can be used to describe motion in a single plane, if the frequency components are weighted and the weighting quoted with the results. These weighted measurements are compared with the frequency levels in the critical wavebands, for each axis, to determine their acceptability. In most practical cases the difference between measuring every 1/3 octave band and an overall weighting, "Is small". But if one is working near the limits of exposure, the more precise method is recommended. As this thesis was not working near the boundary, weighted overall measures could be used.

If vibrations occur in more than one axis simultaneously, the limits for each axis are applied separately. If the magnitudes are similar in each axis, the X and Y components are multiplied by 1.4 and the weighted vector sum of the three axes considered. The effects of the combined motion, in this case, can be greater than in any single axis.

4. RAILWAY RIDE/VIBRATION RESEARCH:

One of the earliest railway ride tests was conducted by British Rail in 1949 (Reeves (date unknown)). The study was done to standardise track quality after nationalisation. A series of tests was carried out using a train and a trolley-bus.

British Rail officials sat in vehicles, facing the walls, as they were accelerated at various speeds. The details of the next curve were announced in advance to the officials, the thirteen occupants then graded the accelerations as:

> 0 = Nothing noticed.
> 1 = Just noticeable.
> 2 = Noticeable.
> 3 = Pronounced.
> 4 = Very pronounced, but not at all uncomfortable.
> 5 = Strong and slightly uncomfortable.
> 6 = Rather uncomfortable.

They concluded that the boundary between comfortable and uncomfortable was 1.22 m/s2 (equivalent to between 6.5 and 7.70° cant deficiency). Sensation number two was selected for the new standard (equivalent to 3.50° cant deficiency).

This research has a number of weaknesses. A group of middle-aged male railway engineers cannot be considered a representative sample. Assessments are likely to be biased due to the respondents' personal characteristics, detailed knowledge and wide experience. The fact that the characteristics of each curve were announced in advance would have made assessments even less representative, as staff would have been able to prepare for any movement. This may also have caused their assessments to be hyper-sensitised.

British Rail later produced curves (Koffman (October 1968)) showing how passenger tolerance varies with vibration frequency and journey length. A ride index (based on these curves) was produced and the value of 3.25 was chosen as acceptable, in both the vertical and horizontal planes. This was later reduced to 2.5 for long distance and 2.8 for commuter stock.

Much of the ride work on railways, in this country, is associated with the development of the Advanced Passenger Train (A.P.T). The aim of the A.P.T. was to increase speeds round curves, without detracting from passenger comfort, by tilting the carriages. Although conventional trains could corner faster without leaving the rails, speeds were limited by passenger comfort and track geometry/maintenance (Reeves (date unknown), Pollard (November 1984), Harborough (May 1986), Chappell (March 1984), Chappell (February 1986), Harborough (July 1986)).

In the A.P.T. tests, trains were run at various speeds above line speed and passenger reactions were gauged. The tests were carried out using British Rail staff (women and older people were under-represented) who knew what they were measuring. Passengers did not, however, know whether they were on a fast or slow run. Most passengers were seated, but some were asked to stand or walk about.

Passengers completed a questionnaire at the end of a trip, where they rated ride between, "Very comfortable" and, "Very uncomfortable". Passengers were asked to press a button, when they considered the lateral ride unacceptable. On average this happened every 2/3 mile. Video cameras were fitted to observe movements in the coaches, but the results were thought to be of little value.

Walking passengers were the most critical of ride, followed by standees and seated passengers. Lateral vibrations were the main problem, though the steady state lateral accelerations produced during a curve, were not. Irregularities had a greater effect when they occurred on curves. Passengers appeared to experience more severe problems when they were not expecting a jolt, as the entrance to a curve was assessed more severely than the exit. No significant differences in response were found, between positions in the coach.

When the train kept to line speeds, 12% of standees were unhappy with the ride, though only 2% of those seated were. Each 10 m.p.h. increase in speed (over the line speed) approximately doubled the number of upset passengers. There were a wide range of perceptions and no, "Jumps" where large groups of passengers became dissatisfied.

The A.P.T's tilting mechanism was found to considerably reduce passenger discomfort. Some errors of perception were evident, as passengers reported improvements in vertical ride when the tilt system was in operation (tilt does not affect this). This suggests passengers have some difficulty in separating lateral and vertical ride.

Finally equations were derived that could predict the proportion of passengers (seated and standing) who would be dissatisfied at any level of engineering ride.

The validity of this research is again challenged by its use of staff as respondents. This sample was also biased towards middle-aged males - although not as badly as earlier research. Asking respondents to hold a button to record discomfort meant they were constantly aware of their assessment role. This may have generated over-critical assessments because of hyper-sensitivity.

5. RAILWAY RESEARCH INTO INTANGIBLE ATTRIBUTES:

5.1. Station Facilities:

One of the earlier attempts to value intangible attributes was made by Vorhees and Associates (March 1971). They were asked to estimate a value for improvements in station facilities at Edinburgh Waverley. This was achieved by asking passengers about today's journey in a structured questionnaire. A hypothetical approach was used and an alternative station was described, as like Marks and Spencers or Euston - but with no structural alterations. Respondents were then shown a photograph of a new train indicator display and plans for the alterations to the station. Interviewees were asked to decide whether they were prepared to pay a premium (a proportion of their current fare) to obtain the improvements. Those who would pay the premium were asked what was the maximum they would pay, for the new facilities. Those who would not pay the premium, were asked if they would pay any less to gain the facilities. Generally, 70% of the interviewees were prepared to pay.

This study was based on current users of the service, who may be happier with the facilities than less frequent users. The method of establishing the financial value of the improvements was also rather crude and could be easily manipulated by the respondent. The researchers also discovered a flaw in their method. Many of the respondents had not experienced the new station at Euston and so were thinking about the old station during the exercise.

Copley et al (date unknown) also considered station investment. In this study facilities were initially considered as a series of attribute levels (for example, covered seating or no station staff). An important feature of this research was its attempt to value intangible attributes. For this to be achieved the levels of such attributes had to be described, this was done using a series of illustrated descriptions.

The relative value of these attributes was established using a priority evaluator. This is a trade-off game, where respondents are allocated pretend money that they can spend on improvements to a station. The improvements that a respondent buys, indicates his relative valuation of the attributes.

In the next stage two bundles of station attribute improvements were traded-off with known attributes, to enable the monetary value of the station's facilities to be established. This was achieved by asking respondents to rank nine hypothetical journeys in order of preference.

It is clear from this study that values can be estimated for intangible attributes, using various forms of trade-off analysis. Although this work has successfully developed levels for intangible attributes, these techniques are unlikely to be useful when valuing ride. Most of the attributes considered by Copley et al are more tangible than ride and thus have some form of measurement that is comprehensible to the public. Station facilities are also more amenable to illustration than ride.

Comfort attributes are often considered to have little effect on the demand for rail travel. But a study (which attempted to model the demand for new stations) in West Yorkshire found that for 30% of rail trips, comfort was the only factor that could explain the decision to use this mode (Preston (1987)).

5.2. General Rolling Stock Attributes:

Research Projects (December 1968) produced one of the first rolling stock studies. They surveyed the London-Portsmouth route, interviewing passengers, businessmen at work and people at home.

On-train respondents were given a list of twelve attributes and asked to state whether they were essential (E), desirable or unnecessary. The results are presented below in terms of importance according to the E rating (approximate figures). Most important were cleaner train toilets E: 42%, followed by newer trains E: 27%, quieter/smoother trains E: 26%, better heating/ventilation E: 25%, more seating E: 23% etc. Respondents were now asked to select only three improvements. These were generally: quieter/smoother trains, newer trains & cleaner train toilets.

In the household survey, interviewers again presented respondents with a series of attributes (which were different from the on-train set). The approximate results were (in order of importance): cheaper fares 58% (not considered on-train), bigger station car parks 40%, newer trains 22% and modernised stations 20%.

In an attempt to produce valuations for rolling stock improvements, Research Projects developed an economic game. Passengers were shown photographs of an alternative train, while the differences were explained by the interviewer. Respondents were then asked if they preferred the alternative. Those who preferred it, were asked whether they would pay a supplement (dependent on the length of their journey, varying between 6d and 2/-) 55% of non-daily users and 25% commuters opted to pay the supplement. Those who would pay a supplement were then asked how much they were prepared to pay.

There was evidence from the interviews that ride quality was a major issue - a number of passengers spontaneously commented on it. Some passengers were actually frightened because of it. When new trains (4-VEP's) were introduced, 80% of passengers had noticed the improvement in ride.

This study was one of the first that attempted to value an intangible attribute. However, its direct approach is more open to abuse by respondents than more recent trade-off designs. It therefore offers little help with the development of the thesis. Although ride appears very important in this study, it should be remembered that trains were much more rougher riding when this survey was done.

M.I.L. Research (1982) conducted interviews in London and the South East, just before the introduction of new Class 455 trains, which significantly improved the service. A household survey, based on rail users, was used because of the complexity of the approach. M.I.L. carried out approximately 2,500 interviews among people aged 16-70, that had used Network South East at least four times in the last year.

M.I.L. listed the improvements wanted by passengers (shown in order of importance). 9% wanted better furnishing or heating, 7% better staff, 7% waiting rooms or toilets, 6% more trains, 6% reduce fares/cost, 5% faster booking, 4% cleaner stations, 3% better catering, 3% more late night trains, 2% better trains, 2% more easily operable doors, 2% more non-smoking, 2% more through-trains and 2% less vandalism and graffiti (8% produced other comments). Ride did not appear. Those who used the train most, were more concerned with quality factors. Passengers were also asked whether they would prefer, improved services at no extra cost (chosen by 61%) or the same service and 5% lower fares (chosen by 35%).

Finally, an attempt was made to value service attributes, using a trade-off game. A number of levels for each attribute and the price of changing them, were specified. Respondents could pay to change any attribute's level. The attributes were divided into two groups, primary and secondary. Primary attributes cost twice as much to change, as secondary ones. The primary attributes were: overcrowding, frequency of service, punctuality, journey time and age of carriages. The secondary attributes were: cleanliness, seating comfort, interchange, cleanliness of windows and information on disruptions. Ride was not included in either set of the attributes.

By noting the proportion of the sample, that were prepared to pay a given percentage of fare to improve an attribute, M.I.L. were able to estimate elasticities for each attribute.

Unfortunately the levels of intangible attributes were poorly described and clearly capable of broad interpretation. For example, the levels of, information on disruptions were: "Very poor", "Poor", "Fair", "Good" and, "Very good". These descriptions are so unrelated to objective measures, that any elasticities produced would be almost impossible to apply.

Steer Davies and Gleave (S.D.G) conducted a study, to establish the best way of replacing the stock then on the London-Bournemouth line (S.D.G. (April 1983)).

The first stage consisted of approximately 100 depth interviews and group discussions, with rail passengers and users of other modes. This established what people considered to be the most significant differences between old and new trains.

In the second stage 200 semi-structured interviews were conducted, both on the London-Bournemouth and London-Swindon lines. Interviews were done on the Swindon line, as this would allow ratings to be compared for old (Bournemouth) and new (Swindon) stock.

Twice as many people were dissatisfied with the Bournemouth service, as were with the Swindon one. Apart from the difference in trains, S.D.G. considered these lines to be similar, but there were many other significant differences. The Swindon line was associated with significantly shorter journey times (using 125 m.p.h. trains, a speed that could never be achieved on the Bournemouth line) and is less congested in the peak. Both these features may account for some of the difference in ratings between the services.

Passengers, on both lines, were asked to compare the two stocks. Photographs were successfully used to distinguish between them. 60% of passengers on each route had experience of the other stock (or an equivalent). Those who did not have this experience, were given descriptions by the interviewer. These descriptions were, unfortunately, often presented in a leading way.

The Bournemouth (4-REP/4-TC) stock was compared to Swindon's Mark 3 coaches in four categories: cleanliness, roughness, noise and newness. The old stock was considered worse in every category, even though there was little measurable difference in cleanliness. Respondents managed to isolate a number of attributes including: upholstery, seat design, noise/ride, compartments and trimmings. Though they found it difficult to specify particular attributes, that caused dissatisfaction with the Bournemouth stock.

A crude attempt was made to estimate the overall value of new stock, by asking passengers whether they would pay more (to have it - Bournemouth, or to keep it - Swindon). 75% were prepared to pay 10% extra and 37% would pay 20% more.

In the final phase 3,000 random postal questionnaires were sent out, along the London-Bournemouth line. People who made more than one trip per year, were asked what they considered to be the best and worst two features of a train trip to London. No account was taken of the strength of these feelings and it may have been useful to ask all respondents about a trip by train, however infrequently they use it, as the perceptions of infrequent users may be different. In this context, it is interesting to note that the postal sample (infrequent users) were slightly more reluctant to pay for improvements, than the on-train sample. Nevertheless the S.D.G. study is one of the few that has attempted to tackle infrequent users.

The best five features of the service were (in order): speed 67%, frequency 40%, the ability to read/work on train 28%, reliability 19% and general comfort 19%. The worst five features were: fares 59%, dirty old trains 31%, travel from Waterloo 18%, travel to station 16%, speed 13% and rough ride 7%.

It is interesting that the ride on the old service, was considered so poor that 7% of postal respondents volunteered comments about it.

M.I.L. Research (April 1984) distributed 3,970 self-completion questionnaires on InterCity trains to and from Euston, Paddington and Kings Cross. They found that 70% of passengers thought the ability to get a seat was the most important feature of a trip. This was followed, a long way behind, by (in order): heating/ventilation, cleanliness, cleanliness of toilets, smoothness of ride and the availability of buffet services. The remainder (in order): luggage space, the availability of a restaurant car, window cleanliness and on-train information were relatively unimportant to passengers. Generally the top positions were maintained across the routes surveyed.

Respondents then rated the attributes of the current service. None of the attributes were considered poor, averaging 3.5 or above, on a five point scale - "Very poor" (1) to, "Very good" (5). Seat availability came top, being considered above good (4.1 - 4.4) on the scale. Luggage, heating/ventilation, buffet, cleanliness and smoothness of ride were considered just below good (3.4 - 4.0). The availability of a restaurant car and window cleanliness were slightly lower (3.3 - 3.6). Toilet cleanliness and on-train information gained the lowest ratings (3.2 - 3.5) though they were still considered favourable.

The features that attracted the most divergent opinions were: seat availability, smoothness of ride, window cleanliness and the availability of a restaurant car. M.I.L. suggested that some of the spread in ride and restaurant ratings, resulted from differences between the services of each region. For example, the Paddington respondents considered ride to be smoother than passengers on other routes - which reflects reality.

M.V.A. (May 1985) conducted a two stage study into passengers' reactions to new trains in the South East. The main aim of this, was to establish a value for new rolling stock. Initially people were approached on stations and general information about them and their journey was collected, they were then invited to participate further. All respondents, whose journeys were greater than four miles, completed a questionnaire at home. Those who were considered suitable after this, then completed a second questionnaire.

Respondents were asked to rank nine alternative journeys in order of preference. The journeys were defined by four factors: journey time, cost, reliability and whether the stock was, "Old" or, "New". From the choices made, the value of each factor could be estimated. Reliability was found to be very important - twenty times more than new stock. Over the whole sample the average value for new stock was 9% of fare - with a confidence interval of (2.3 - 16.1%).

Respondents were next asked to assess the importance of eleven attributes and the performance of the old and new stock with each attribute. Respondents were reminded of the trains with photographs. People could find it difficult to recall the old stock, even though M.V.A. had chosen survey locations where new stock had been introduced within the last 18 months.

Each attribute was rated as: "Very important" (1), "Fairly important" (2) or, "Not important" (3). The performance of each train was rated as: "Very good" (1), "Good" (2), "Average" (3), "Poor" (4), "Very bad" (5). Both scales were treated as simple intervals, allowing average scores to be produced. These are presented in the tables 2.1 and 2.2.

TABLE 2.1: IMPORTANCE OF ATTRIBUTES (IN ORDER).

Ventilation in Summer	1.26
Clean Carriages	1.30
Good Chance of a Seat	1.30
Heating in Winter	1.39
Support While Standing	1.50
Ease of Getting On/Off	1.59
Comfortable Seats	1.61
Smooth Ride	1.63
Open Spacious Carriages	1.90
Luggage	2.19
Colour Scheme	2.52

The new stock was generally preferred to the old, except for ventilation in summer, seat comfort and support while standing. New stock was synonymous with: open/spacious carriages, cleanliness, comfortable seats, smooth ride, ease of getting on/off, layout, decor/colour. But for almost everyone: cost, frequency, reliability & getting a seat were more important. Virtually no difference was found with luggage and the probability of getting a seat.

TABLE 2.2: PERFORMANCE OF OLD AND NEW STOCK
(BEST PERFORMING STOCK UNDERLINED).

	NEW	OLD
Colour Scheme	2.30	3.23
Smooth Ride	1.78	3.15
Clean Carriages	2.43	3.43
Heating in Winter	1.96	2.71
Ease of Getting On/Off	2.46	2.85
Luggage	2.76	2.75
Ventilation in Summer	3.12	2.49
Comfortable Seats	2.74	2.61
Support While Standing	3.29	2.92
Open Spacious Carriages	2.14	3.38
Good Chance of Seat	2.56	2.51

When disaggregating the results, M.V.A. found that peak travellers were more concerned with crowding and ventilation. Older people were more interested in: smoothness of ride, warmth, comfort of seats and spaciousness. The elderly, surprisingly, gave less priority to, standing support. Women showed more concern for the train environment, than men.

This research shows ride to be relatively unimportant, even relative to other comfort attributes, like heating. The removal of passengers making very short journeys has biased the sample and may have affected the results. Some of the attribute levels presented in the trade-off were ambiguous, for example, "Your train is quite often more than five minutes late". It is not easy to attach an objective measure of punctuality to such a statement and the findings are therefore difficult to apply. Describing the alternative trains as, "Old and, "New" is emotive and could have generated an excessive valuation of stock.

Nevertheless, this is one of the few rail studies which has conducted choice exercises at home, in the the appropriate decision environment (which can have a big effect, see chapter 3). M.V.A. have also used a more advanced (trade-off) approach for estimating values than previous rolling stock research.

M.V.A. (January 1988) recently carried out investigative interviews with passengers on comfort and other attributes. Ride was included under the, "Comfort" heading, which included a small number of negative and positive statements about ride.

Headings were scored according to the proportion of statements that were negative, for business, commuter and leisure trips. Comfort had 39-49% negative statements (one of the lowest). Other scores were: toilets 92-100%, maintenance and safety 90-100%, information 80-91%, cramped 72-87%, cleanliness of stations 72-82%, cleanliness of trains 71-81%, catering 69-75%, price 59-81%, network coverage 62-80%, staff 44-70%, general appearance 35-69%, reliability 36-59%, speed and journey time 9-33%, miscellaneous 9-33% and quality of mode 0-6%. The most mentioned heading (negative and positive) was reliability, followed by cleanliness and comfort.

Within comfort, comments were made on: disabled facilities, luggage space, heating, seats, ride, lighting, ventilation, smoking, noise, doorways and tables. Generally, in the comfort heading, the most frequent comments were about ride. Three times more people liked the ride than disliked it (the Class 455 units in the survey, are good in this respect). There was a slight difference in priorities between groups, with commuters complaining more about overcrowding.

Further information on passengers' views is available from the consumer bodies that represent them. Their findings are available in a series of reports (Central Transport

Consultative Committee (June 1987 and June 1988), Transport Users Consultative Committee (September 1987)). Information is also available from letters in journals (for example, Modern Railways and Rail) or newspapers. It should be noted that such information is biased, as it mainly represents the views of passengers who complain.

These sources suggest, that there is some concern with the new trains being introduced. There are a number of design features the public are unhappy with, though ride does not appear to be one of them. The problems are (in no particular order): lack of luggage space - especially for cycles and pushchairs, obstructed views - often caused by seats not aligning with windows, reduced legroom, narrow seats, overcrowding, lack of toilets, inadequate ventilation and a noisy environment. Such problems are particularly apparent in the Provincial sector.

Finally information is available from, "Monitor" surveys conducted by the business sectors at regular intervals. Network South East's six monthly monitor suggests that passengers are, in general, satisfied with ride.

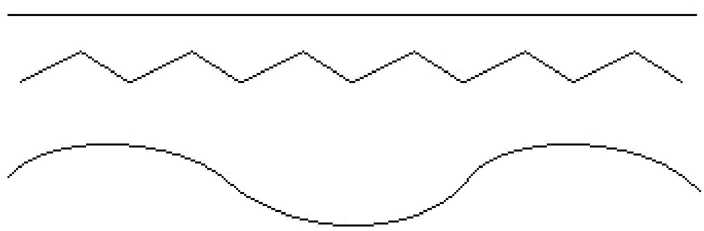
5.3. Provincial Rolling Stock Attributes:

Steer, Davies and Gleave (August 1982) conducted a study to see how passengers reacted to a new railbus, on a rural railway line in Wales. Ride quality was measured on a scale of: "Very Smooth", "Fairly Smooth", "Neither", "Fairly Rough" and, "Very Rough".

Attribute ratings were established, both for the stock then being used (1950's-60's generation diesel multiple units - D.M.U's) and the Class 140 railbus. S.D.G. noted a fatigue effect, as the ride (and some other) ratings deteriorated with distance.

One innovation of this research was the use of, "Squiglograms" to further investigate the effects of ride quality. These are pictorial representations of the kind of ride experienced by the passenger. Respondents were asked to select the one that most effectively portrayed the movements they felt.

FIGURE 2.3: EXAMPLES OF SQUIGLOGRAMS.



Squiglogram responses give some idea of whether the ride was characterised by jolts from jointed track, or long wave vertical motion caused by suspension damping. Unfortunately, such techniques would be of little use when trying to value ride, as the pictures cannot be effectively related to engineering measures.

Respondents considered the railbus smoother while changing gear, accelerating and decelerating. It was also considered smoother on continuous track. But the D.M.U's smoother ride over rail joints, more than compensated for this and was therefore considered to offer the better overall ride.

Steer, Davies and Gleave et al (November 1981 and January 1983) have conducted other research to compare various forms of Provincial rolling stock. These investigations used the same rating scales, as the Welsh research, allowing comparisons to be made.

S.D.G. compared three types of train: Class 140 railbuses (including the prototypes: L.E.V. and R.3), conventional D.M.U's and the Class 210 (a new more costly conventional unit). Investigating these results should provide some interesting insights into people's perceptions of ride.

TABLE 2.3: RIDE RATINGS FOR ALL STOCK (%).

UNIT	VERY SMOOTH	FAIRLY SMOOTH	NEITHER	FAIRLY ROUGH	VERY ROUGH
R.3. (5)	18	52	7	18	3
L.E.V. (4)	6	41	9	28	7
140 (1)	11	41	15	24	5
140 (2)	5	42	15	28	8
140 (3)	4	17	10	56	10
D.M.U. (1)	3	44	12	31	7
D.M.U. (2)	3	53	18	21	4
D.M.U. (3)	6	42	10	35	6
D.M.U. (6)	3	41	15	34	3
210 (6)	47	41	5	6	1

MISSING VALUES ACCOUNT FOR THE REST

(1) Birmingham - Stratford on Avon.
(2) Preston - Colne.
(3) Swansea - Shrewsbury.
(4) Unspecified Route.
(5) Unspecified Route.
(6) Unspecified Route.

S.D.G. found that the initial reaction to the ride of the railbuses was more severe and that opinions mellowed, as experience of the vehicle increased. It is clear from table 2.3, that a vehicle's ride is significantly affected by the track. On poor track the D.M.U. was preferred to the Class 140; but as track quality improved the Class 140's ratings increased further than those of the D.M.U. So on high quality track, the ride of the Class 140 was preferred to that of the D.M.U.

The Class 210 unit was very popular with passengers, they often likened it to an InterCity 125. This appeal is illustrated, by the proportion of respondents that were satisfied with ride quality and noise levels - compared to the old D.M.U.

TABLE 2.4: PROPORTION OF PASSENGERS SATISFIED WITH RIDE AND NOISE (%).

ATTRIBUTE	210	D.M.U.

RIDE	88%	44%
NOISE	80%	22%

Although the overall ratings for a given train on each track section are different (this may not be significant), a cross-sectional study was used and so we cannot be sure that individuals could distinguish between the ride of each track section.

According to the, "Squiglograms" the railbus types were characterised by a notchy ride (35-40% passengers chose such an option). In fact the Class 140 was commented on as having, "Square wheels" - illustrating the problems of railbuses on jointed track. The D.M.U's were considered to have more of a wave motion (60-70% passengers). The most popular response for the Class 210 was a straight line (44%), though 37% thought it had an oscillatory motion. Only 9% of respondents considered the D.M.U. ride to be like a straight line.

The A.B. Group Ltd. (June 1984) also investigated passengers' attitudes to new Provincial trains. Ride was described to respondents as, "Motion". This research failed to find a significant difference, between the mean ride scores for old D.M.U's and the Class 141 railbus. These means were calculated from ordinal data and so the result, although plausible, is not really valid.

Advertising Principles Ltd (November 1986) conducted further research, into the effects of the introduction of an enhanced railbus (Class 143) in the North East. They produced a three phase study. Phase one considered the old D.M.U's, phase two was done immediately after the introduction of the Class 143's and phase three was done several months after this.

The change in attribute ratings, after the introduction of the new trains, indicates the size of, "Newness effects". These occur when newly introduced products achieve higher than expected ratings, because of a kind of novelty value. A Small deterioration in attribute scores was noticed in phase three.

The Class 143's out-performed the old D.M.U's in all areas, apart from luggage space. However, there was still considered room for improvement with: ride, seats, ventilation and noise.

The ride quality ratings of all three phases are shown in table 2.5. Again, "Motion" was also used to describe ride quality.

TABLE 2.5: MOTION SCORES (%)

RATING	PHASE 1 D.M.U.	PHASE 2 143	PHASE 3 143
V.Good	8	24	16
Q.Good	34	45	44
Neither	26	20	24
Q.Poor	20	8	10
V.Poor	12	3	5

Hudson, Payne and Iddiols (August 1987) compared a more conventional replacement train (Class 150) with the old D.M.U's. They found that the Class 150's were perceived to have a smoother ride than old D.M.U's.

The final piece of research concerns an attempt by Steer, Davies and Gleave (March 1984) to estimate the passenger reaction to an experimental passenger coach.

A new body, designed by British Leyland, was placed on an early 1960's British Rail Mark 1 underframe (with B4 bogies). This experimental coach was then placed in a rake of early 1970's Mark 2f air conditioned coaches (also mounted on B4 bogies). As both types of coach used the same bogies, engineering ride should be similar. Apart from this, there were few similarities between the coaches.

Approximately 160 people were interviewed in each type of coach, both groups had similar personal characteristics. Respondents were asked to compare attributes between the Leyland and Mark 2f coaches.

Perceptions of ride were measured, using both ratings and Squiglograms. Ride ratings were not significantly different between coaches, even though many other attributes were rated differently. Oscillatory movements were the most common response for both types of coach. The fact that ride perceptions were the same for both types of coach, suggests that Halo, Newness and contamination effects were minimal.

The similarity of ride assessments is particularly significant: as the Leyland coach (which had opening windows) was perceived to be significantly noisier, than the air conditioned Mark 2f's. Despite the close association of ride and noise, respondents appear to have isolated the two.

6. THE PREVIOUS ATTEMPT TO VALUE INTANGIBLE ATTRIBUTES:

6.1. General:

The M.V.A. (May 1986) InterCity rolling stock study, is the only work that has attempted to estimate the values of intangible attributes, including ride and is thus worth more detailed attention.

6.2. Priority Evaluator:

Valuations were produced for the individual attributes of a stock improvement package. This was done using a game, called a priority evaluator (P.E). Each respondent was shown a board, on which a series of attributes were listed - at various levels. The cost of achieving each level was also shown.

The number of attributes on the board, was limited by what respondents were able to consider in a single experiment. It was felt that large errors would result, from doing separate experiments for each attribute. The attributes presented were: heating and ventilation, quality of seating, internal layout, on-train information, decor, cleanliness, toilet facilities, doors, luggage facilities, staff and ride quality.

Four levels of each attribute were presented. The levels were chosen on the basis of what appeared to be just noticeable to interviewees (during the preliminary interviews) and what were (according to British Rail) feasible investment options. Pictures for the extreme levels on the P.E. board were included to make the task easier.

Ride was one of the most expensive elements to improve. The exact ride descriptions and costs were:

Level 1. Rough ride with frequent jerks, sufficient to spill drinks from a cup. Cost: 0.

Level 2. Generally smooth ride, but occasional jerks sufficient to spill drinks from a cup. Cost: 2.

Level 3. Ride quality smooth, but handwriting not very easy due to vibration. Cost: 6.

Level 4. Very smooth ride, almost like home or office. Cost: 10.

Respondents were asked to identify the existing level of each attribute on the board. They were given a budget, that they could use, to buy improvements listed on the board (indicating the importance they attached to each attribute).

The budgets given to respondents, were designed so that they could only reach the highest levels in some attributes. This ensured that respondents had to trade-off improvements. Respondents were given 15 or 25 units (moderate and high investment levels) depending on the exercise. They did not have to exhaust their budget, as this may have caused strange allocations.

6.3. Preliminary Findings:

Wide variations were found in the perception of current attribute levels, even within the same groups of stock and individuals. These differences may be partly due to maintenance, but it is most likely to result from varying perceptions.

Current ride quality was generally considered to be, in the middle two levels. The mean assessed level of ride, for all passengers, was 2.4.

TABLE 2.6: PERCEPTIONS OF RIDE LEVELS ON MARK 3 COACHES.

CLASS	LEVEL ONE	LEVEL TWO	LEVEL THREE	LEVEL FOUR
First	14	53	42	5
Standard	17	94	69	8

Passengers were interviewed on Mark 3 coaches, some of which had been refurbished. M.V.A. identified Halo and Newness effects. A Halo effect occurs, when improvements in various attributes complement each other. This means that passenger satisfaction is greater for the whole package, than the sum of the individual attributes. The Newness effect inflates the benefits attributable to new stock during its early life.

Refurbished carriages tended to received more favourable ratings. For example, passengers reported a slightly smoother ride on the refurbished stock, even though all coaches used the same bogies. It is suggested some of this difference may be the result of the Newness and Halo effects. M.V.A. suggested two other possible causes. The refurbished coaches may have had more recently serviced bogies - though no relationship was found. The improved levels of at-seat service on the refurbished trains would have reduced passengers' need to move about.

M.V.A. asked respondents to value attribute levels that were not related to a particular train, in an attempt to contain Halo and Newness effects.

Quality of ride was identified as an important aspect of stock quality and one with which respondents were not satisfied. Passengers tended to associate this with the track, rather than the rolling stock. It was found that the better travelled passengers, were able to distinguish between the ride levels of the routes they used. Respondents considered the Great Western Main Line the smoothest, followed by the East Coast Main Line and the West Coast Main Line. This order corresponds with engineering opinions expressed in, for example, Modern Railways. The ability of passengers to make such comparisons, could be of value later in the thesis.

Respondents felt that ride quality should be improved to give greater comfort and to allow activities not currently possible. A minority found it intolerably difficult to drink or walk at present. Sway was seen as more of a problem than vibration or judder. It was suggested, that at-seat service would reduce the need for passengers to move around and therefore contain the effects of ride. First Class passengers were keenest to improve ride quality, or (if this was not possible) at-seat service. Standard class passengers seemed to reserve their budget for other attributes like: seat quality, layout and facilities. Most respondents were appreciative of recent improvements, realising the great cost involved.

Noise and quality of ride were said to acquire importance as the trip progressed. The extent of this importance, was partly dependent on the activities planned during the trip. Business travellers, in particular, found when writing that, "Sway and vibration made this activity difficult". Other passengers were not generally disturbed by movement, apart from sudden jerks. The latter movements were only really seen to be a problem, when moving or drinking.

6.4. Utility Model:

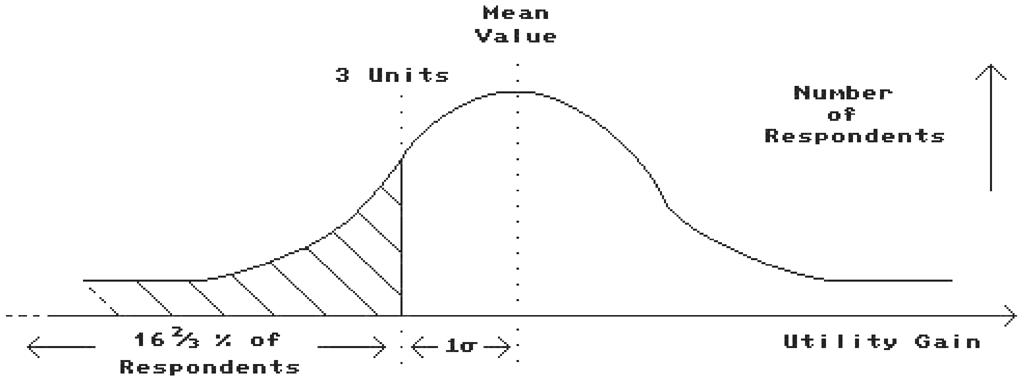
To determine the value of each attribute, M.V.A. developed a method that could establish the change in utility that results from a movement between any attribute levels. This method is explained below, using ride quality as an example.

An individual who believes ride quality to be at level one, has the option (subject to his budget) of moving to any higher level. Such a movement will increase that individual's utility. But, before this individual decides to allocate any of his budget to this improvement, he has to consider the utility he might gain by spending the units on something else (opportunity cost). For example, an improvement in noise levels may yield more satisfaction than the better ride. So only a proportion of the sample (who gain more utility from the better ride, than they lose by not spending the units on something else) will choose to buy an improvement in ride.

M.V.A. assumed that the utility gained from a series of improvements in any attribute, is additive. Further, respondents will not all experience the same gain in utility, from a given movement. In the absence of any better information, M.V.A. assumed that the distribution of individual utility gains (resulting from an attribute improvement) was normal.

To obtain an overall value for the utility gained from a movement between an attribute's levels, the mean gain in utility has to be estimated.

FIGURE 2.4: ESTABLISHING THE MEAN GAIN IN UTILITY.



To illustrate the procedure, we assume that ride is currently considered to be at level one. We want to establish the mean gain in utility for a change from level one to level two. Suppose 83 $^1/_3$% of respondents have chosen to move from level one to level two, at a cost of (say) three units. We know that these respondents will have gained at least three units worth of utility from this move. It follows that 16 $^2/_3$% of respondents, would have gained less than 3 units worth of utility and so have chosen not to allocate any of their budget to this improvement.

We can show this position on a graph - with utility on the X axis and the number of respondents on the Y axis. This graph shows the distribution of respondents' values for the change in ride between level one and level two. M.V.A. have assumed that this distribution is normal. The mean value for this change in ride will therefore lie at the centre (with 50% either side), it is this value that we want to establish.

From a table of normal curve areas: we can see that the 16 2/3% of respondents with ride values below three units, are located to the left of one standard deviation below the mean. We can therefore say that the mean ride value is one standard deviation above three units. If we can establish the size of this standard deviation, we can estimate the mean ride value. M.V.A. assumed that the standard deviation was equal to the mean divided by two.

In our example the mean can now be established by solving the equations below:

Standard Deviation = Mean - 3
Standard Deviation = $^{\text{Mean}}/_2$

Mean - 3 = $^{\text{Mean}}/_2$	(Multiply by two)
2 x Mean - 6 = Mean	(Subtract mean)
Mean -6 = 0	(Add 6)
Mean = 6	

A large standard deviation was chosen, to reflect the respondents' broad perceptions of current attribute levels. M.V.A. also wanted to ensure that only a small portion of the distribution would be negative. The chosen value, implied that 2.5% of respondents would derive negative utility from any upward movement. M.V.A. produced a sensitivity analysis for this value and found little variation as a result of changes to their assumption.

When a respondent had selected two rolling stock improvement packages (moderate and high investment) these were valued in a further trade-off exercise. Respondents ranked nine cards containing customised levels of fare and journey time, as well as the stock improvement packages. Once this procedure was completed, the value of each change in an attribute could be stated.

M.V.A. expressed these attribute values as a, "% of fare". For example, if the value of a given change in ride was 3% of fare, a passenger paying £10 for his ticket would (on average) consider this change worth thirty pence. This change in ride would therefore have the same effect on patronage as a 3% change in fares.

Standard errors are provided for the valuation estimates, but they do not include the error from the stated preference experiments. The values obtained for the overall investment packages are shown below.

TABLE 2.7: VALUES OF COMPLETE ROLLING STOCK PACKAGES.

INVESTMENT LEVEL	FIRST CLASS VALUE (% OF FARE)	STANDARD CLASS VALUE (% OF FARE)
Moderate	7.9 +/- 2.9	8.0 +/- 1.6
High	11.1 +/- 3.7	14.5 +/- 3.4

Although the values, in percentage of fares are similar for First and Standard Class passengers - First Class fares are higher, producing a result one would expect. As each individual is valuing their ideal package, these figures probably overestimate the value of, "Real" new stock. From comparing the average expenditure on an attribute with that of each individual, M.V.A. recommend a write-down of 80% on these figures.

The estimated values of the ride change between each level, are shown below.

TABLE 2.8: RIDE VALUES (% OF FARE).

CLASS	1-2	2-3	3-4
(Moderate Investment)			
First	1.13	2.55	3.46
Standard	0.99	2.33	3.69
(High Investment)			
First	2.23	2.85	3.87
Standard	1.04	2.31	3.64

6.5. Critique:

M.V.A's was one of the first attempts to value intangible product attributes. One would therefore expect some weaknesses in the method.

It could be argued, that as trade-off exercises do not present respondents with a decision whether to travel, the results may be invalid (for example, Louviere (January 1988)). The method used by M.V.A. (and many others) infers the choices respondents would make, from the results of the trade-off games.

Asking respondents to value abstract attribute levels has contained Halo and Newness effects. But as these levels were not related to specific trains and therefore individuals' experiences, respondents may have had difficulty visualising the levels they were supposed to be valuing (for example, Kroes et al, Fowkes et al, Bradley, Hensher et al (January 1988) and Green et al (September 1978)). There is clearly a trade-off here, between the errors caused by Halo and Newness effects and those resulting from inaccurate interpretation of attribute levels.

The assumptions used to establish mean utility values for moves between an attribute's levels are open to question. Such assumptions would not be necessary, if a simpler form of trade-off (for example, ranking or rating a series of attribute combinations) had been used. But, sensitivity analysis showed changes in assumptions to have little effect. Also the need to value large number of intangible attributes, would have made the implementation of a simple trade-off difficult.

Perhaps the biggest single weakness with the research is that the attribute levels, valued by respondents, bear little resemblance to technical measurements. This is especially the case with ride. The vagueness of the four ride descriptions, means that it has been very difficult to apply M.V.A's ride values. Even where the results have been applied, the values have to be considered very cautiously. So despite the care was taken to produce accurate results, their practical use is limited.

The overall values of rolling stock produced by the research are also flawed, as each individual was valuing a different package. These values can therefore only be used as a broad guide.

Finally M.V.A's values were generated for InterCity and may not be applicable in a Provincial environment.

Although there are problems with M.V.A's findings, this was only a first attempt in a difficult area. The M.V.A. research does provide some useful results, giving ride values that can be usefully compared with any figures produced later in the thesis. A number of insights are also provided by M.V.A, that should make any future investigation easier and more accurate.

The weaknesses of M.V.A's research and the need for a Provincial value of ride, mean that further research is required into intangible attributes.

7. CURRENT PRACTICE:

The ride descriptions valued by M.V.A. have been related (as far as possible) to levels of engineering ride via, "A perceived ride index". The vagueness of M.V.A's descriptions mean that they are associated with a wide range of possible engineering levels. British Rail have chosen engineering levels that suggest the maximum value of ride (Frederick (October 1987)).

Previous research suggests that attribute values on Provincial services are lower than those on InterCity services (British Railways Board (June 1986)). This means that M.V.A's ride values are arbitrarily reduced (by the same order as other attributes' values) when applied to Provincial services. The uncertainty involved in the application of these ride values shows the need for further research.

8. CONCLUSIONS:

8.1. General:

Much of the work on ride quality is designed to find the tolerance limits of humans, these boundaries usually range from discomfort to those affecting health. Some research has attempted to compare engineering measures of ride with perceptions. A summary of the results is available in the various standards (British Standards (1987), International Standards Organisation (1985), International Standards Organisation (1985)).

Only one study has gone a stage further (M.V.A. (May 1986)) and tried to identify how levels of ride quality affect the demand for travel - the main objective of this thesis. The indirect relevance of much of the previous work means that it can only provide pointers for the design of the thesis. Nevertheless by reading broadly, a picture has been built that should enable the research to more effectively progress.

8.2. Pointers from Previous Research:

The following issues have been highlighted by the review of previous work in ride and related areas.

It is difficult to combine vibrations into a single ride figure: so a series of values have to be reported to describe engineering ride. When considering the ride of railway vehicles, many of the accelerations have been found to be so small that they can be ignored. Those omitted are pitch, yaw and often longitudinal acceleration. Roll is observed infrequently and generally only with work on high speed curving. The ride of the vehicle can mean that roll and lateral acceleration have similar effects on passengers. The availability of equipment and low significance of longitudinal acceleration and roll in previous work, means that this work will use only lateral and vertical acceleration measures.

Assessments of ride can be affected by a number of factors. These are listed below. Hyper-sensitivity and the use of (experienced) staff volunteers can produce over critical results. People produce widely varying ride assessments (partly due to vague scaling). Women appear to more affected by vibration and generally the older one gets the less effect vibration has. There may be a fatigue effect, causing ratings to deteriorate over longer journeys.

The wide interpersonal variations reported in this review (for example, Cooper et al (1978) and M.V.A. (May 1985)) mean that sampling has to be carefully controlled, if ride values are to be representative of the population.

Passengers whose activities are interfered with by ride (for example, writing or walking) are likely to produce more severe ratings. Passengers who are able to anticipate movements are less likely to be affected by ride (making the modern designs of multiple unit, with no forward view, appear questionable).

Noise could be a serious contaminant in ride quality ratings. M.V.A. have identified the Halo and Newness effects, which may also interfere with ratings. It appears that people can isolate intangible attributes, though there is no agreement on the level of interference from other attributes. Only a few studies have produced results that give direct indications of Halo and contamination effects - these were found to be negligible. However, a small Newness effect has been identified. Despite the ambiguity of these results, care will have to be taken to minimise such effects.

Early attempts to establish values for intangible attributes used fairly crude techniques, taking little account of non-engineering research. People were often just asked, how much they would pay for a change. This is clearly open to abuse by the respondent, as it quite obvious what the research is attempting to discover. Such an approach is also further removed from an actual choice situation, than the conjoint techniques used in marketing.

The descriptions of attribute levels, used in much of the previous work, are vague and so any values that have been obtained are difficult to apply. For example ride levels have been described, as rough and smooth. M.V.A. made some attempt to improve the scaling of ride perceptions by mentioning physical consequences, like the spilling of drinks. But clearly a more precise descriptive scale has to be developed, that can be related to engineering measures.

A number of studies have asked respondents to compare different trains, that they have used. It appears that passengers are able to remember journeys, on other trains in the past and photographs have been successfully used to differentiate between them. People seem to be able to distinguish ride between trains, though not necessarily between track sections.

There is some disagreement, between surveys, on the importance of ride quality. It appears that the importance of an attribute, may be related to the levels currently experienced by the passenger. Those studies that found ride to be unimportant were done on InterCity coaches and Class 455 units - on which the ride is good. The ride experienced by those respondents who considered it important, was significantly worse.

M.V.A. have established a series of ride values from interviewing passengers. But, there are indications that infrequent users may have lower values for such improvements. M.V.A's values have other weaknesses and so this research attempts to produce a method, that will generate more valid intangible attribute values.

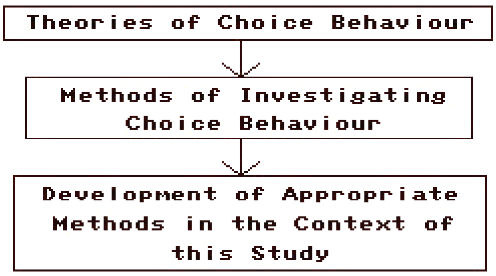
Chapter Three Developing the Research

1. INTRODUCTION:

The main aim of this chapter is to develop a series of practical approaches for estimating the value of a change in ride quality.

The main choice theories developed in previous marketing and psychology studies are described. The ways in which choice behaviour can be investigated are then considered. The application of these methods in the environment of this study are then considered. From this, all the possible ways of establishing a ride value are outlined and assessed on the basis of the knowledge gleaned from previous work (described in chapter two). Those approaches that are clearly impractical, or unlikely to produce significant results, are not to be given further consideration.

FIGURE 3.1: DEVELOPMENT OF APPROACHES.



The approaches that survive the initial evaluation are discussed in slightly more detail. It becomes clear from this, that a preliminary investigation (described in the next chapter) is necessary before the remaining techniques can be developed further. The issues to be considered in this further investigation are discussed.

2. THEORIES OF CHOICE BEHAVIOUR:

Foxall (1983) describes two schools of choice behaviour. The first group are the cognitive theorists who argue that people make decisions in a conscious calculated way. The other group are the behaviourists who believe that decisions are made in a less ordered way.

2.1. Cognitive Theories:

Cognitive theories are more conventional. It is argued that everyday experiences cause individuals to be constantly bombarded with information. An individual's ability to recognise such information via their senses is greater than their ability to process it effectively. To avoid overload, this information is rapidly processed in short term memory using a series of filters which cut out unnecessary information.

Information passing through individuals' filters is processed, generating over time, a series of beliefs and consequently attitudes - with regard to themselves, issues and objects. Attitudes can be defined as, "A learned predisposition to behave towards an object in a given way" (Tuck (1976)). An example may be, "I hate travelling to work by train". The development of these attitudes may in turn modify an individual's views of the relevance of any future information and will thus determine the shape of the filtering system in the future.

Fishbein has produced one of the most renowned cognitive theories. This is outlined below, both as an example and also because it provides some useful pointers for later research in this thesis.

One of the main cognitive theories of choice behaviour is that of Fishbein. His research suggests a series of relationships between beliefs, attitudes and behaviour. Fishbein has developed techniques for eliciting the most influential or, "Salient" beliefs behind an individual's choice process (Tuck (1976), Towriss (July 1981), Towriss (January 1984), Foxall (1983)).

Fishbein states that actual behavior is strongly related to intended behaviour; though the greater the gap between establishing behavioural intentions and the act, the weaker the relationship between them.

Behaviour $\approx$ Behavioural Intentions

Behavioural intentions are a function of attitudes to the act and normative beliefs (what other people think about the act). It is important to measure attitudes to the act and not attitudes to the object, as the latter is not specific enough - producing inaccurate results.

BI = W1 ATTACT + W2 NB

BI	Behavioural Intention.
W1, W2	Weights.
NB	Normative beliefs.
ATTACT	The attitude towards the act.

Attitudes are said to be a function of beliefs about the act. Beliefs have two components: an evaluative aspect and the strength (probability) in which the belief is held. A belief might be, "Going to work by train is slow".

ATTACT = Σ { B E }

B	Strength of the Belief about the act.
E	How favourable, or not, the belief is.

The final part of the equation is concerned with normative beliefs. An example might be, "My wife does not like me using the train to get to work". Normative beliefs are also made up of two elements: the belief itself and the desire to comply with it. This part of the equation has been found to add little to the model and so is often left out.

NB = Σ { SNB MC }

SNB	Strength and direction of normative belief.
MC	Motivation to comply with normative belief.

People appear unable to manipulate more than five to nine beliefs when making a decision. Consumers have been been found, when making complex decisions, to use an attribute to cut-off certain alternatives to simplify the process. The attribute chosen will be one where there is a great difference between the choices (Bither and Klein (September 1987)).

Fishbein argues that individuals will concentrate on the beliefs that are most important to them, in distinguishing between choices. These salient beliefs should be the most readily elicited. To achieve this, the respondent is presented with an open question about a specific act like, "What comes readily to mind, when thinking about going to work by train?".

2.2. Behavioural Theories:

Foxall (1983) describes an alternative explanation of choice behaviour known as, "Behaviourism". This school believe that man is not a rational decision making animal, as Fishbein et al indicate. Behaviourists therefore argue that there is no point in studying conscious choice processes.

Behaviourists suggest that beliefs and attitudes are formed by behaviour and not vice-versa. The experience of past decisions will therefore generate decisions in the future. They suggest that individuals are creatures of habit and only change their behaviour when something forces them to. If this was the case, the existence of an improved rail service (with a smoother ride) would not affect a car user - until his vehicle became unavailable and he was forced to try the new service. This person now has experience of the new rail service and may change his attitude to it.

Behaviourists believe that the decision environment has an important effect on purchase decisions, this has been shown by, for example, Ginter and Miller (February 1979). An individual's beliefs and attitudes will change with the environment. It is therefore important to measure the elements of a decision in the place where the decision is made. Behaviourists suggest that only where very situation specific measures are used, is there much of a relationship between attitudes and behaviour.

The link between the decision environment and purchase decisions, means that great care must be taken when designing hypothetical choice situations. Fowkes and Wardman (January 1988)) state that, "As the values increasingly diverge from individuals' experiences or from what appears plausible, the stated preference responses can be expected to become less reliable". As individuals have, "Widely different experiences" an attempt should be made to produce a customised design to minimise any error.

Cognitists have conventionally assumed that an individual conducts a search to find a product that solves a problem. But search costs will mean that not every possibility is investigated. Behaviourists argue, that in reality very few are. Foxall (1983) states, that it has been consistently found, that many consumers have little regard for information made available to assist in the choice process: habit or unconscious decision making, is thought to control many choice situations. Behaviourists argue that a simple everyday decision like, whether to avoid a small puddle, would not be the result of some rational decision making process. Cognitists would argue that such a decision is made, perhaps almost unconsciously, before the puddle is reached.

2.3. Overview:

Cognitist theories (like the Fishbein model) have provided an understanding of the way people make choices, successfully relating beliefs, attitude and behaviour. But to be able to make predictions of aggregate behaviour, these theories have to be developed one stage further. This is generally done using microeconomic utility theory. Such approaches have been able to estimate values for intangible attributes and are described in detail later in this chapter and the following ones.

Neither of the two theories are without weakness. For example, Cohen and Miniard (February 1979) suggest that there is a flaw with the Fishbein approach: as the normative element in the model is not really separable from the individual's attitudes.

The behaviourist view is also contradicted by some evidence. It is implausible to suggest, that an enhanced rail service would only attract additional users who were forced to try it out. This may apply to some individuals, but the introduction of such services in the past (British Railways Board (June 1986)) have shown effects that are too large to be explained by this alone.

Although Foxall (1983) describes the two explanations of choice behaviour as rivals, closer examination makes this dichotomy less apparent. Foxall admits that there may be two types of choice: low involvement where the outcomes are not dramatic and little mental effort is undertaken and high involvement, where there may be a conscious choice process.

It could also be possible (when faced with a recurring decision, like a journey to work) that a cognitive choice is made at the beginning and the outcome then becomes habitual, until something happens that forces a change. However Foxall states that such splits are not yet supported by evidence.

Although the behaviourist approach may provide some new insights into the choice process, it is more of a critique than a theory and it is therefore difficult to use it predictively. Even if the decision process is not conscious, as argued by the behaviourists, it may still be rational. This means that it can be modelled, using similar techniques to those successfully developed - based on a cognitive hypotheses.

3. INVESTIGATING CHOICE BEHAVIOUR:

3.1. General:

On the basis of the discussion in the previous section, it is clear that any attempt to examine and predict choice behaviour is best based on a cognitive hypothesis. This can be related to microeconomic utility theory (for example, Laidler (1981)) which is conventionally used to model choice behaviour. The microeconomic theory is now outlined.

Any activity undertaken by an individual is associated with some level of satisfaction (utility). It is hypothesised that individuals are rational and will thus choose the course of action that maximises this utility (or minimises disutility). The demand for any product is thus related to the amount of utility individuals associate with its consumption.

Transport is unusual as it is not usually demanded for its own sake, it is generally required by individuals so that they can take part in some activity (for example, to reach work, go shopping or get to the cinema). Transport can thus be said to have a derived demand as it is a means to an end. The cost of transport in time, money etc. is likely to reduce the utility associated with the activity it is used to reach. It is therefore important to consider the net amount of utility gained from both the travel and the activity at its end. If this net utility is below the net utilities associated with other activities, the individual will not demand the trip.

As an example consider an individual with three options: staying at home, a journey to work by bus and a journey to work by train, these yield five, ten and fifteen utils respectively. We would expect (if there were no other alternatives) this individual to take the train to work.

It is argued that an individual does not know the exact level of utility they will gain from pursuing a choice: they therefore make an estimate of the consequences (Tuck (1976)). This estimate is based on a subjective probability distribution and is the level of utility the individual expects to gain from choosing an alternative (expectancy value model). This subjective probability element can be related to the strength in which beliefs about the action are held.

Choice of Action = f [UO1, UO2, UO3... UOn]

UOx = Expected Utility associated with Option X.

No investigation into consumer choice is likely to correctly estimate the utility any individual associates with a particular course of action. We cannot obtain all the information on which such choices are based and so our predictions will be generated from imperfect information. It is therefore not possible to precisely predict an individual's behaviour.

When trying to predict choices, this problem of imperfect information is tackled using random utility theory (for example, Ben-Akiva and Lerman (1985), Bates (January 1988), Fowkes and Wardman (January 1988)). According to this theory, the utility associated with any choice is made up of the component we are able to determine from any investigation and a random element.

UOx = Dx + ex

Dx = Determined Utility.
ex = Error Term - Random Element.

The choice between options is therefore specified in terms of probability. The assumptions made about the distribution of this error term determine the form of model that is developed. For example, "A Weibull distribution yields the most commonly used form of random utility model, the multinomial logit model" (Fowkes and Wardman (January 1988)).

Any good or service can be looked at as a combination of the attributes that make it up. For example a particular rail service may be considered: quick, expensive, difficult to reach and comfortable.

Consumers are believed to begin a choice process by identifying a need. This need can be satisfied by certain product attributes. For example, an individual may feel a need to get to work faster. The performance of each product with respect to attributes associated with this need is determined through a search and learning process. Consumers then make judgements of the merit of having attributes at certain levels. This information is integrated to form overall impressions of various products (Louviere (1988)).

The utility an individual associates with any good or service, can thus be broken down into a series of part utilities for each attribute. It is the relative sizes of these part utilities that determine an attributes value. If one of these utilities represents the money cost of the option, we are able to state the value of any attribute in financial terms.

Utility of Option = f [Ua1, Ua2, Ua3... Uan]

Uax = Utility associated with attribute x.

This exact form of this model varies between studies, for example part utilities may be added or multiplied. Additive forms are generally the most popular (Kroes and Sheldon (January 1988)).

The two main groups of techniques that can be used to estimate the coefficients of the attributes in such a model are now described.

3.2. Revealed Preference Techniques:

With revealed preference techniques the choices individuals make in real situations are observed. Attribute levels for each possible choice are also noted. This information allows estimates of the utilities associated with options and attribute levels to be established. This is usually done using mathematical techniques, like regression or maximum likelihood.

This approach is best illustrated with an example. Assume competing train and coach services, the train costs £2 and the coach £1. The train takes half an hour and the coach one hour. If an individual chooses the train, he reveals that his value of journey time is greater than £2 an hour.

Unfortunately the procedure is rarely this simple. In our example the train may have been more comfortable and had more luggage space than the coach - our consumer may have spent some of the extra money on these features. Isolating attributes from each other is one of the major difficulties with the approach. For a revealed preference approach to be successful we must therefore find a situation where the attribute we wish to study is one of the few that varies.

These issues are discussed in more detail later in this chapter, where the implementation of a series of revealed preference approaches is considered.

3.3. Stated Preference Techniques:

3.3.1. GENERAL:

Revealed preference techniques have traditionally been used in the study of demand. However in recent years stated preference techniques have evolved to a point where they are also capable of producing predictive models of choice behaviour.

Heeler et al (February 1979) and Brinberg et al (March 1986) list a number of ways of measuring attribute importance - though not necessarily values. This can be done through: Open ended elicitation, trade-off analysis, importance ratings, subjective probability measures, paired comparisons, or information search. These techniques are outlined below. Although they are described separately, there have a number of common features and some research has combined aspects of each.

3.3.2. OPEN ENDED ELICITATION:

Open ended elicitation, involves respondents stating the attributes that come readily to mind, while thinking about a choice decision. The ease with which the attributes are extracted, gives an indication of their importance in the choice process. This is characterised by the Fishbein approach, described earlier.

But Fishbein's technique has to be applied very specifically for it to succeed. It is also very difficult to make aggregate predictions of behaviour on the basis of these attitudes and beliefs (Tuck (1976), Foxall (1983)). Furthermore, as the Fishbein technique only identifies the, "Salient" or most important beliefs about an action, the effect of intangible attributes (like ride) may be too subtle to be picked up. However the Fishbein approach may still be useful in indicating how salient ride is.

3.3.3. TRADE-OFF ANALYSIS:

Trade-off analysis estimates values for attributes by asking respondents to rate certain products characterised by different levels of specified attributes. To compare these alternatives respondents have to trade one attribute for another. For example, a choice may have to be made between a fast expensive car and a slow cheap car, the respondent thus has to trade speed with cost. Best and McCullough (February 1979) argue that trade-off analysis generally, is stable over time and reliable. These techniques are described in more detail later (chapter eight) where a trade-off model is developed.

3.3.4. IMPORTANCE RATINGS:

With importance ratings, respondents are directly asked to rate a series of attributes on some consistent scale. It is suggested that techniques, like this, where respondents do not have to trade attributes can lead to excessively high ratings (for example, Guilford (1954)).

3.3.5. SUBJECTIVE PROBABILITY MEASURES:

Subjective probability measures establish the importance of attributes by asking respondents the probability of them choosing a product, while varying the levels of each attribute individually. The size of each attributes' effect on the probability of choosing the product, indicates the importance of the attribute. An adaptation of this technique may be able to generate attribute values.

3.3.6. PAIRED COMPARISONS:

With a paired comparison approach, subjects are presented with pairs of attributes and asked to indicate which of the two would be more important to them, in evaluating some product. This is done for all possible pairs of attributes, allowing estimates of importance to be made. But again this approach does not provide values for attributes and the production of individual estimates is not possible with this technique (Ackerman et al (March 1986)).

3.3.7. INFORMATION SEARCH:

One of the main information search approaches, uses the Information Display Board. With this technique respondents are asked to make a hypothetical choice: they then take information about the products from a display board until they have enough to make a choice. By doing this respondents reveal the attributes that were most important to them while making the decision.

3.4. Technique Conclusions:

From the previous discussion it is clear that the most effective way to produce a value for a change in ride quality, would be to develop techniques based on the standard microeconomic utility model. For any rail journey ride quality would be one of the attributes represented in such a model. This thesis has to find a way of isolating the utility associated with various levels of this attribute. By expressing these utilities relative to those of money, we are able to produce a financial value of a ride change.

There are two broad groups of techniques that can be used to isolate the utility associated with changes in levels of ride quality, these are used in the next section to generate a series of possible approaches. After the initial review it would appear that stated preference techniques are most likely to be successful.

Although there are many such techniques for measuring the importance of product attributes, importance may not indicate value. There are a number of reasons for this. The relationship between importance and value may not be linear (for example, Tuck (1976)). When making independent assessments people often rate all attributes as important. Trade-off analysis or an adaptation of subjective probability measures offer the most potential for establishing a ride value - allowing predictions of aggregate behaviour to be made.

Other techniques may also be of use in the research. An open elicitation approach, such as that developed by Fishbein, could help establish the importance of ride in the early stages of the work.

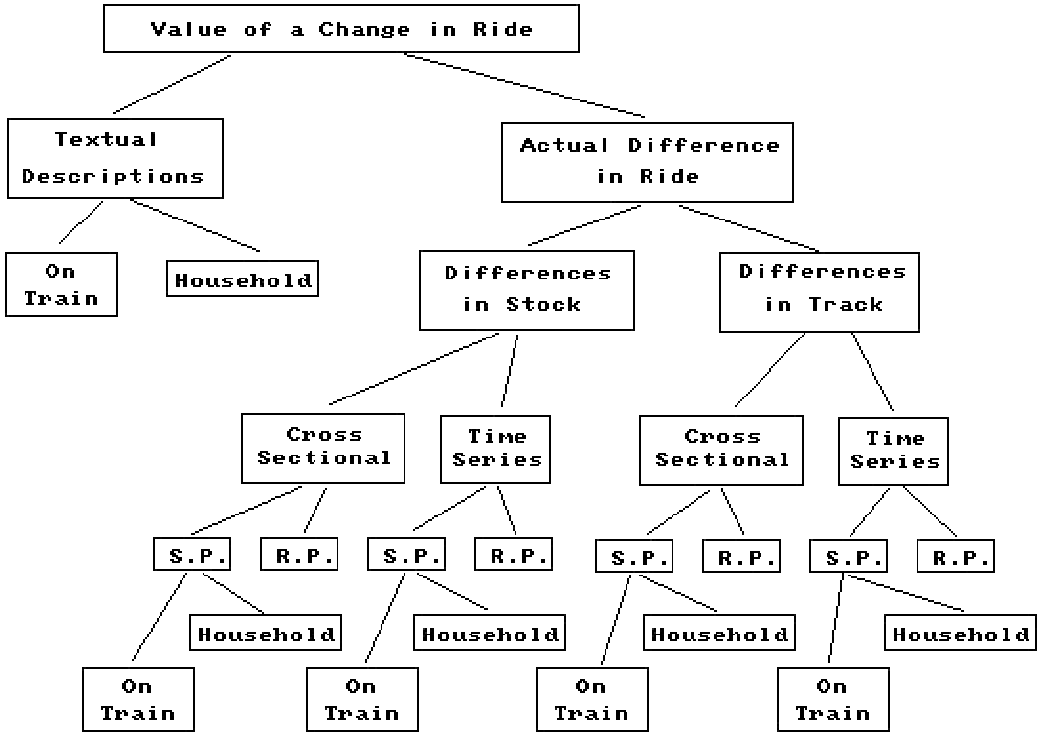
It is also clear from Fishbein's work, that respondents can only handle a very limited number of attributes (between five and nine) when making a choice decision. If such a choice process is to be simulated later in the research, even fewer attributes could be used, as respondents will have less motivation than in a real situation.

Situational characteristics appear to have a very important effect when measuring the decision process. It has been found that behavioural predictions are less reliable, when data is collected in an inappropriate environment.

4. POSSIBLE APPROACHES:

The ways of producing a valuation of ride can be looked at as a hierarchy (see figure 3.2).

FIGURE 3.2: POSSIBLE VALUATION APPROACHES.



To establish a value for a change in ride quality, a series of people must consider the difference between two levels of engineering ride. They must then value this change in terms of something - probably money. As mentioned in chapter one the major difficulty with this procedure, is that engineering measures of ride are meaningless to the general public. We cannot therefore present people with engineering measures of ride and expect them to place a sensible value on the difference between them.

Respondents must be presented with descriptions of ride that they can relate to, based on their experience of travel. These descriptions should be linked to technical scales, so that they act as a proxy for engineering ride measures. The way in which these proxy measures are presented and the method used to value them are the main elements of any ride valuation approach.

Respondents can be presented with two basic forms of alternative ride measurement. Firstly they can be presented with actual levels they have directly experienced. For example, "Your train journey from London to Bedford last week". Alternatively they can be presented with textual descriptions of ride levels based on their general travel experience. For example, "The ride is so rough that your drink spills frequently".

Looking at the actual changes in ride experienced by respondents, reveals a further division of techniques. From the previous chapters it is clear that, the ride felt by passengers results from the roughness of the track and the way that a vehicle compensates for the this roughness. Ride quality is therefore influenced by the quality of the track and the quality of a vehicle's suspension. Other factors can also affect the sensations felt by a passenger, but these are of less importance. Changes in either track or suspension, could make an attempt at ride valuation possible.

Directly experienced ride levels can be compared for one service over time (time-series) or between a number of services at a point in time (cross-sectional).

The effects of changes in ride quality can be measured using either stated or revealed preference techniques. Stated preference approaches can be implemented as an on-train or home based survey.

The groups of techniques that result from this discussion are now considered in turn. An approach based on textual descriptions of ride is developed first. Approaches based on ride levels experienced with different stock are then considered. Finally, valuing differences in ride produced by varying track quality are discussed.

5. TEXTUAL SCALE DIFFERENCE IN RIDE:

This approach would deal with the ride scale problem by developing a series of textual ride descriptions, which are comprehensible to respondents and related to measures of engineering ride. Respondents would be asked to value movements along this alternative scale using stated preference techniques. Changes in ride quality would be traded-off by respondents against known attributes like fare or journey time.

A textual ride scale must fulfill two main requirements. Firstly it must be directly related to engineering measures of ride. If this is not so, we will be unable to determine the levels of engineering ride that respondents have produced a value for. This will make the results of limited practical value - this is precisely the problem experienced with M.V.A's value of ride (M.V.A. (May 1985)).

Secondly the textual scale must be easily comprehended by the general public. If the scale is not clearly understood, respondents will not be sure what levels of engineering ride they are valuing. If respondents have difficulty interpreting the scale, they may not take the task seriously - producing meaningless results. They may even fail to complete the exercise - reducing the response rate. Respondents may also make a best guess. Any such guesses are likely to be more varied than the respondents' real values of a change in ride. Presenting respondents with a poorly constructed textual ride scale is likely to result in a lack of precision that will reduce the practical value of the findings.

Initially it seems that the best way to generate an effective textual scale, would be to consider events objectively associated with particular levels of engineering ride. Such a scale could be based on: injury, sickness, inability to read, inability to write, or the spilling of drinks. The most effective of these events would appear to be the spilling of drinks, as this could be directly related to accelerations on a train.

Such a scale could use a series of objective statements, such as, one spill per hour etc. Spilling of drinks is likely to be something previously experienced by the majority of passengers. But, the drinks phenomenon is only a representation of peak values of ride and cannot be easily related to an average (R.M.S) measure. So even with the drinks measure, which initially appears the most effective, there would be only a loose relationship between engineering ride and the textual scale.

This textual scale approach is similar in a number of ways to the approach used by M.V.A, which is the only study to have produced a value of ride. It is therefore wise to give this approach further consideration. But before proceeding any further with this approach, it is important that we establish a suitable form of language for the textual scale. This is best done using a depth interview (described in the next chapter). Langauge could be investigated by asking, for example, "What is meant by ride?" or, "How rough is this service?".

This approach could be implemented as an on-train or home based exercise. At this stage both will be considered for further development, as the technique can operate in both environments with little alteration. The effects of the choice of survey location (for example, sampling and the decision environment) will be considered later.

6. ACTUAL CHANGE IN RIDE - DIFFERENCES IN STOCK:

With this group of methods respondents are asked to value differences between actual ride experiences, associated with various forms of rolling stock. These differences in engineering ride would result from the different suspensions employed on the rolling stock. Suspensions can differ both between and within, types of rolling stock.

6.1. Different Suspensions Alone:

Ideally we would be dealing with a series of similar trains each with different bogies, thus offering various levels of engineering ride. In such a case, the only differences between the trains would be ride quality. Any valuation of the differences between trains would therefore be almost solely due to ride quality. Unfortunately this situation rarely occurs naturally. In the past a number of similar Mark 1 coaches were fitted with bogies of very different riding characteristics (for example, Glover (1987)). Unfortunately few of these coaches still exist - especially with the older rough-riding B1 bogey.

In an ideal world two coaches of the same multiple unit would be modified, so each offered a different level of engineering ride. Passengers could then be quizzed about the differences between them. If the same train was run over a particular route for any stretch of time, it may even be possible to observe effects on the seating positions of regular travellers. Such passengers could be asked if there was any reason for them selecting their seat. Each coach could be painted slightly differently, so that respondents could distinguish between them.

But such an approach is clearly not practical with a Ph.D. thesis. The idea of finding two similar trains with different riding characteristics therefore has to be abandoned.

6.2. Different Rolling Stock:

In this case respondents would be considering the difference in ride between a number of different types of train. This would mean that ride would not be the only difference between the trains. The ride effect therefore has to be isolated from the effects of all the other changes in attributes.

6.2.1. REVEALED PREFERENCE:

A revealed preference approach would mean acquiring patronage figures for a number of origin-destinations, served by stock with different levels of ride. Each origin-destination's patronage levels could be related to the level of ride experienced, using some form of mathematical model. This would allow the value of ride to be estimated.

Unfortunately there is one unsurmountable problem with this approach. Ride will not be the only difference between trains. Those with the best ride will tend to be of a newer design, most other attributes will also have improved in the new trains. This means that all the independent variables, representing stock attributes, will vary together - resulting in perfect multicollinearity (for example, Groebner and Shannon (1985)). Generally this perfect multicollinearity (singularity) makes it impossible to isolate the ride effect from those of other attributes. For example, if ride was the only attribute included in the model it would be greatly over-valued, as it would act as a proxy for all the other improvements.

The near impossibility of producing a value for ride with such an approach, means that it will not be given further consideration.

6.2.2. STATED PREFERENCE - TIME-SERIES:

This approach would be implemented after the introduction of new trains (with a better level of ride quality) on a line. The most effective way of developing an such an approach, would be to consider the effect on passengers of a reduction in ride quality to its original level.

The effect of the ride improvement, resulting from the introduction of the new stock, could be estimated. But this would be considerably more difficult than estimating the value of a reduction to the original level. For example, estimating the value of the ride improvement, would mean establishing the number of journeys made by our sample, before the new stock was introduced. We would then have to isolate the proportion of the change in trip rates that was due to ride, rather than other improvements. The isolation of ride, in such a general case, would be extremely difficult. Such an approach is likely to suffer from major recall errors, as it is unlikely that travellers can accurately remember their trip rates of up to six months ago.

Predicting a value for a reduction in ride quality is more appropriate for what British Rail Provincial have in mind. By implementing the approach after the ride improvement, a group of previously infrequent users would be included in the sample. Finding a representation of such people, when considering the value of a ride improvement, would be very difficult. No attempt will therefore be made, to estimate the value of a ride improvement with this approach.

This group of approaches tackle the ride scaling problem, by referring to the ride levels associated with certain trains. People will be selected on the new trains, who have experience of the previous trains. People who do not have experience of the previous stock would be excluded from the survey, as such passengers could not give a realistic valuation of the change in ride. All respondents thus have experience of two engineering ride levels.

Interviewees would be asked to think back to when their decision to travel was made; hypothetically giving them time to plan for alternatives (Harrell (1987)). Passengers would then be asked to re-make this decision assuming the previous levels of ride quality. The proportion of current passengers who would no longer make today's trip would allow an estimate to be made of the value of ride.

Engineering ride would be measured on the route before and after the introduction of new stock. This allows the stated change in patronage to be related to a change in engineering ride.

This technique could be applied as a household survey - asking a respondent about his last or next trip. Fishbein states, that to effectively investigate the way people make decisions, we should obtain information just before the actual decision (for example, Tuck (1976)). Considering a respondent's last trip is therefore less likely to reveal choice processes, than interviewing just before or during that trip.

The best household approach would be to ask a respondent about his next trip. But, the low trip rates on rural lines, would make it very difficult (in a household survey) to find enough people who were planning their next trip. For example, Centre for Transport Studies (1989) found that 68.2% of passengers used the train less than once a week. This household approach will therefore not proceed.

This technique assumes, that all passengers expected to travel on the new train and had a good idea of what it was like, at the time of the decision. Implementation is therefore likely to be most effective, if it is done some time after the introduction of the new stock. However the greater the lag after introduction, the greater the error in recalling the ride of the old stock.

Despite these problems it is believed that this approach has sufficient potential for it to be developed further.

6.2.3. STATED PREFERENCE - CROSS-SECTIONAL:

This approach would again mean finding respondents who have experience of a number of types of stock. The ride scaling problem is again tackled by referring to the ride of actual trains on a specified line. But in this case, a line would be selected where both trains were running together. This means that the ride on both trains will be fresh in respondents' minds.

With this approach respondents have not been able to assume that their current trip would be on the new stock. So we have to estimate the effect on respondents, of the differences in ride quality, using a hypothetical trip. To ensure an accurate response this hypothetical trip should be as close to reality as possible. A value for the change in ride would be produced by asking respondents to choose between a series of imaginary trips characterised by different levels of ride and other attributes.

As a hypothetical trip is being considered this approach could be implemented on-train or at home. The issues involved in this decision are discussed more fully in chapter five.

7. ACTUAL CHANGE IN RIDE - DIFFERENCES IN TRACK:

This group of approaches are based on respondents having experienced a number of levels of ride, as a result of differences in track quality.

7.1. Time-Series:

These approaches rely on finding a line where track quality has recently been changed, resulting in a different level of ride quality. In the past this has often occurred as a result of laying continuously welded rail (for example, Cooper (1984)). Respondents could compare the current ride with their memories of the previous ride and indicate the value they place on the difference.

But, there are a number of problems with such approaches. As the old situation has been eliminated, it will be difficult to establish the size of the change in engineering ride that respondents are valuing. Previous levels would probably have to be estimated, by looking at other un-modernised lines.

For these approaches to be effective, the change in track quality would have to happen quickly to reduce recall error. But, it is rare for large sections of a route to be modified in a short period and such a situation is very unlikely to be found during this research. If implementation was delayed until the full replacement of track, recall error is likely to be severe. Intermediate experiences of ride would also interfere with respondents' memories of the original ride level (Baddeley (1982)). If the study were done before the full replacement of track, the difference in engineering ride over a particular journey would be small.

Changes in track quality are often associated with corresponding changes in journey times. Overall the difference in ride experienced by the passenger before and after the change in track quality is therefore likely to be small.

These time-series approaches will therefore not be developed further.

7.2. Cross-Sectional:

This approach would mean finding a series of origin-destinations characterised by different standards of track and thus ride quality.

7.2.1. REVEALED PREFERENCE:

With a revealed preference approach we would need to compare levels of patronage for a series of origin-destinations. By applying mathematical modelling techniques we could then estimate the effect of differences in ride quality on these patronage levels.

This form of revealed preference technique is likely to be more successful than that discussed previously (which considered the effect of ride differences between trains). But, there are unfortunately still difficulties, even with this, more promising revealed preference technique.

Variations in ride, produced by differences in track quality, are likely to have a very small effect on patronage. This means that there will be insufficient co-variation in ride and patronage, for an effect to be isolated using revealed preference techniques (for example, Kroes and Sheldon (January 1988)). Even if the services on each route are similar, there will be differences in other factors which have a far greater effect on patronage. For example, population size and characteristics, location of stations, and unexplained effects. The Transport and Road Research Laboratory (1980) state that such difficulties will result in, "Unreliable estimates of their individual effects so that little confidence can be placed in forecasts of the effects of changes in them".

These more important differences would have to be specified in any model, or the ride variable may act as a proxy for their effects. Any model would therefore need a substantial number of independent variables. A large sample of origin-destinations would have to be considered, to explain the effects of all the independent variables.

At the start of this research Provincial patronage data was still patchy, despite the introduction of improved computerised ticketing procedures, for example PORTIS. It may therefore have been difficult to establish flows for a large number of origin-destinations.

A revealed preference model, in such circumstances, is unlikely to produce a usable value of ride. Even if it did, the large amount of data required would make this an inefficient way of proceeding. None of the previous work in this area, reviewed in chapter two, was based on revealed preference techniques. Nash (1982) states that, "For many aspects of quality which are difficult to measure (comfort, cleanliness, convenience of exact timings) market research techniques are almost inevitably chosen". This revealed preference approach is therefore not be developed.

7.2.2. STATED PREFERENCE:

There are two ways of developing this technique. We could ask respondents to compare the ride quality of various trips they had made, with the same stock, on different track sections. Alternatively we could find a long distance route comprising of a variety of track standards. Respondents could then be asked to compare sections of their current trip. The former approach could be implemented as a household survey.

Both these approaches rely on respondents having experience of a number of ride levels in the same stock. Respondents' perceptions of ride and the corresponding engineering levels, could be measured throughout the routes. The value respondents attached to the differences in ride, that they had experienced, could then be estimated. This estimation would probably be made using some form of trade-off task.

8. METHODS WORTHY OF FURTHER INVESTIGATION:

The previous discussion has greatly reduced the number of possible techniques for establishing a value of ride. Only the following groups of approaches are considered worth developing later in the research:

a). Textual Scale Change in Ride: On-train and Household.

b). Actual Change in Ride - Stock Change: Stated Preference - Time Series - On-train.

c). Actual Change in Ride - Stock Change: Stated Preference - Cross-Sectional - On-train and Household.

d). Actual Change in Ride - Difference in Track: Stated Preference - Cross-Sectional - On-train and Household.

9. BASIC STATED PREFERENCE ISSUES:

Although all revealed preference techniques have been rejected (as they were clearly incapable of producing useful results) this is not to say that the preferred stated preference approaches are flawless. They are however more likely to yield some form of ride value.

There are a number of issues that have to be considered with these stated preference approaches. As the research develops the weaknesses of these approaches will be discussed in more detail and remedies for these shortcomings discussed. Some of these issues would be usefully investigated in a preliminary interview, these are considered below.

We have already seen from the discussion of scaling methods for the textual scale approach, that the form of language used by members of public (as opposed to engineers) to describe ride and its effects has to be studied. This will allow respondents to be presented with tasks that are comprehensible and unambiguous.

It is important to discover how sensitive peoples' perceptions of ride are and to what extent they notice changes in ride. This is especially important when considering differences in ride produced by the track, as the worst track will be traversed at lower speeds - evening out differences in ride.

Interference effects have to be investigated as they could affect respondents' ability to isolate ride from other attributes. These effects can take three forms, "Newness", "Halo" and, "Contamination" effects. There is a good deal of overlap between them and each author appears to have his own definition.

The Halo effect causes the ratings of individual attributes to be coloured by an overall impression of un/favourableness. The Newness effect occurs because a new train has a sort of novelty value, increasing the ratings of attributes. Contamination occurs where respondents are unable to completely isolate one attribute from another - this is expected between ride and noise. If it was found that respondents have difficulty in isolating ride from other attributes, those approaches based on individuals' experiences of ride could be invalidated. These effects are discussed in more detail in chapter five.

It is important, with all stock based approaches, to find unemotive ways of referring to old and new trains. A way has to be found of describing previous levels of ride, without suggesting they were inferior. These descriptions must also ensure, that respondents know which stock they are being asked about.

People may become hyper-sensitised if they know we are interested in ride. This may cause people, who did not consider the change in ride to be important, to attach an inflated value to it. One way of countering this may be to ask passengers, after a change, whether they detected a difference in any quality of service attribute (so as not to lead them into a belief that ride had changed). If no ride difference was mentioned, its value could be assumed to be zero.

Assuming passengers who have not detected a change in ride have a value of zero may seem severe. It could be argued that ride may subconsciously be an important element in the change. However, the work of Fishbein et al (for example, Tuck (1976)) suggests that individuals can report the salient beliefs that determine their choice decisions. Even if there was this subconscious element, it is unlikely that an individual who has not noticed a change in ride could accurately value it. One problem with this is that, as the travel decision was made in the past, ride may have been salient and is no longer (or vice-versa).

10. ISSUES TO BE CONSIDERED IN THE PRELIMINARY INVESTIGATION:

To proceed with the approaches selected in this chapter, there are a number of areas that need to be investigated. Initially this is done using unstructured depth interviews (developed in the next chapter). Further issues are considered later, allowing the methodology to be developed in more detail.

The issues to be considered in the preliminary interviews are listed below:

a). To establish what ride quality means to the public and to generate an effective descriptive scale.

b). To investigate how sensitive people are to changes in engineering ride.

c). To consider the operation and scale of interference effects. To investigate ways of minimising Newness, Halo and Contamination effects.

d). To find ways of referring to the ride of old and new stock without using emotive language.

e). To investigate the effects of hyper-sensitivity. To consider whether a change in ride is significant, if it is not mentioned by a respondent.

Chapter Four General Investigation

1. INTRODUCTION:

This chapter describes the development of the initial investigation into the passengers' responses to levels of ride quality. As noted in the previous chapters, in order to satisfy the objectives of this thesis it is first necessary to conduct a general investigation into the ride issue. A number of issues have to be considered at this stage to enable the effective development of the work. These issues were defined at the end of the previous chapter and are chiefly concerned with the form of language used by the public to describe ride and gauging their sensitivity to any changes. The results of the interviews and their implications for the research, are discussed.

The interviews were conducted on the London-Poole, Network South East, line during July and August 1988. Although this is not a rural line, there were a number of reasons for conducting the interviews on it. New stock had been recently introduced that Provincial considered to be a considerable improvement for the passenger, part of this was a significantly better ride. This overall improvement would allow an investigation of peoples' ability to isolate the change in ride from that of other attributes. The layout of such trains and the long journeys being undertaken made this service more suitable for depth interviewing than many Provincial lines. The London-Poole line was also easily accessible.

2. CHOICE OF DEPTH INTERVIEW TECHNIQUE:

This exercise was conducted in the form of a depth interview. This is an interview with no rigid format and could almost be considered a guided conversation. There are a limited number of set questions and prompts to which the interviewer can add to investigate a issue in detail.

Depth interviews were chosen for the initial research as they are the most effective way of establishing a general picture of the issues involved. They also allow any interesting leads to be followed up (for example, Social and Community Planning Research (1972), Applied Psychology Unit (1988), Transport and Road Research Laboratory (1980), Open University (1979)).

By using such a general format respondents are able to talk around the full range of the subject and thus provide insights that the interviewer may have missed, during the desk based study. The Transport and Road Research Laboratory state that, "They offer a deeper understanding of travel behaviour in individual circumstances and, to the extent to which the understanding can be generalised, it may provide pointers to the ways in which present empirical models can be improved".

By carrying out the depth interviews at this early stage in the research, it is still possible to alter the later research design, as a result of any new insights.

3. SAMPLE CHARACTERISTICS:

A sample size of thirty-three was obtained, which corresponds with that suggested in the literature (for example, Social and Community Planning Research (1972)). The small sample and the possibility of variation across people (see chapter two) meant that a quota sample was used to ensure a representation of various types of people. This, it was hoped, would allow some attempt at identifying differences between types of people. The sample size and application of such quotas means that the overall sample may not be representative of the population of potential rail users.

The sample consisted of (approximately) one quarter first class and three-quarters second class, within this a 50:50 male:female split was chosen. The age breakdown was done equally on the basis of: young (under 30), middle aged (30-50), and senior (50 plus). Age was not asked - so this is an approximation.

TABLE 4.1: SAMPLE STRUCTURE.

First Class	9	Standard Class	24
Male	5	Male	12
Under 30	0	Under 30	3
30-50	3	30-50	6
50 plus	2	50 plus	3
Female	4	Female	12
Under 30	2	Under 30	4
30-50	2	30-50	5
50 plus	0	50 plus	3

The interviews were carried out over one week and were mostly conducted on off-peak services. Generally semi-fast services were caught to Waterloo and fast services back - following the pattern of working of the new stock. Passengers were approached on the train, if they fitted the sample requirement and were sitting alone in a group of four seats. Lone travellers were selected as it was their particular language that was wanted, not that of a group. However on two occasions groups, of three travellers, were interviewed together in the form of a mini group discussion. The elimination of groups travelling together could have biased the results, though this is considered unlikely, as the sample was spread fairly evenly over different categories of people.

It is important to stress that the sample size is small and if broken down into sub-samples we are dealing with very small numbers. It should also be noted that, as with all on-train surveys, frequent travellers are over-represented. Less frequent users would only be picked up in a household survey and they may have different attitudes towards rail.

When asked about coach as an alternative, passengers were very dismissive, this may not be the case with a more realistic representation of the population. So to increase patronage, the peculiar characteristics of the infrequent user may have to be investigated. There was very little evidence of people making extra trips as a result of the new trains. This could be the result of: only some of the new stock being in service and, for passengers west of Bournemouth, the problems of changing at Bournemouth.

Bias may have also resulted from too little coverage of peak passengers, as it was not possible to sit next to respondents and record an interview. Some interviews were conducted on peak trains after they had thinned out and a number of commuters were interviewed, who happened to be leaving work early. Nevertheless the sample is biased towards the off-peak user.

4. APPROACH AND INTERVIEWER INTRODUCTION:

The interviewer's lack of experience of depth interviews, meant that some M.Sc. students were used to pilot the whole procedure. A number of changes were made as a result of the pilots. These consisted of five full interviews and approximately twenty discussions with students etc, about the wording of the more difficult questions.

During the interview, notes were made on a checklist form. The interview was also tape-recorded unless the respondent appeared unhappy with this arrangement (this only happened once). Once the reasons for the recording, and its confidential nature, were explained there was usually no objection to a tape-recording. Recording the interview ensured that people were not misinterpreted and that any bias could be investigated by an independent source. It also proved impossible to write down everything that was said, while keeping the flow of the interview going. Finally the recording was found valuable when interpreting the rough notes made during the interview. Listening to the recordings showed little evidence of leading questions or any other problems.

It was agreed with N.S.E. that late trains would be avoided, to ensure that passengers' responses were unbiased. In practice this never happened, though one train was missed out as it was replaced by old stock at the last minute. The interviews took approximately ten minutes to complete, though it also took about ten minutes (sometimes more) to find a candidate that fitted the sample. After the interview, respondents usually had some general comments to make and so they were wound-down, usually taking five minutes, after the interview.

The interviewer was introduced to passengers as discussing, the trains on this line. This introduction may have caused some people to concentrate too much on local issues, when asked about British Rail in general. However the local angle did appear to make people more interested. The original introduction was more detailed, but proved to be too much of a mouthful on a busy train. It was found most effective to give a short introduction and then expand if the respondent wished to know more before committing themselves.

Ride was not mentioned in the introduction so that all language was developed independently by the respondent. As the interviews could be long, it was important that sufficient time was available in which to interview passengers. This, convenience and the current mixture of old and new stock, made the Poole line the choice for the depth interviews.

The interviews took place on the new stock - so that respondents were able to think about the changes that had occurred. The interview started with a filter question, to eliminate people who do not have experience of the previous trains. It turned out that virtually everybody had.

To ensure that the respondent was thinking about the same previous trains as the interviewer, photographs were used as a reminder. These photographs showed the inside and out of an empty 4-CIG unit (appendix two). The coach had the new, "Bright stripey" seating material and although not strictly identical to the 4-REPS and 4-TC's previously used on the line, the differences were not noticeable to passengers. Seven people were using the new trains for the first time and so their statements are the result of first impressions. These impressions may be more or less favourable than more experienced travellers views.

The interviewer followed a list of questions and prompts in a specific order, using the same language for each respondent. The interviewee was encouraged to talk, until he began to repeat himself or wandered off the subject. The next question/prompt was then issued. The interviews flowed so well that some respondents were giving information required by later questions before being asked. Following questions therefore had to be asked carefully, to avoid giving the impression that previous answers were ignored. Any interesting comments were followed up - the checklist therefore represented the minimum form of interview. A brief description of the interviewee was written to put the statements into context.

It was important that passengers realised what the context of each question was. The questions were generally about their whole experience of train travel, rather than just the trip they were on. As mentioned previously, a number of passengers jumped straight into the local scene - partly because of the wording of the introduction and also because it was the only line they had experience of. This did cause some problems, as passengers were not clear when asked about the effects of ride, whether they were being asked about the new trains, the old trains, or generally. Some of these problems could have been resolved by piloting the interviews on-train instead of on students. But such a pilot would be very difficult to arrange and as this was only a preliminary interview, the student pilots were felt to be a reasonable compromise.

5. ELEMENTS ELICITED:

The following elements were studied in the depth interviews:

5.1 Language:

This covered the words used to describe ride (for example, bumpiness etc). Care was taken to produce a term that covered all elements of ride - vertical and lateral accelerations. The relationship between noise and ride was investigated. An unemotive way of referring to old and new stock was also looked for.

5.2. Sensitivity to and Isolation of Ride:

People were asked to talk about the changes with the new stock. This was intended to give some indication of the respondents ability to separate out the various attributes. It also told us whether the change in ride was noticed.

5.3. Other Information:

Some more general information was elicited, which was given to N.S.E. It was also expected that passengers would bring up some points, that had been overlooked in the desk study.

6. THE INTERVIEW CHECKLIST AND LOGIC BEHIND IT:

S.1 introduced the interviewer to the respondent and reassured the latter of the genuineness and confidentiality of the approach.

FIGURE 4.1: DEPTH INTERVIEW CHECKLIST.

S> represents a statement by the interviewer.
Q> represents a question.
P> represents a prompt.

S.1> I am sorry to trouble you. I am doing research on the trains on this line for my Ph.D. Could I please record a ten minute interview with you?

Q.1> This type of train is replacing that shown in the photograph. When did you last use one of the trains that are being replaced? <If never, end interview>

Q.2> Thinking about your experience of rail travel in Britain. What, if any, features are you are unhappy with?

S.2> From now on think only about your journeys on this particular line.

Q.3> What, if any, changes have you noticed in the quality of your journeys with the introduction of these trains?

P.1> Anything else? <Repeat until nothing else>

Q.4> Which, if any, of these changes might make you more or less likely to use the train?

P.2> <Read the changes back that they have noticed>

Q.5> What do you understand by the phrase, comfort of the train? <Show card>

Q.6> What do you understand by the phrase, being moved around by the motion of the train? <Show card>

Q.7> Describe the effects, if any, this being moved around has on your journeys.

P.3> Some people have problems reading or drinking on a train.

Q.8> How big a change, if any, would you say there has been in the ____See Q.6____ of a journey with the introduction of these trains?

S.3> Thank you very much for your help.

Age: ______ Sex: M F Class: 1st Standard

Q.1 acted as a filter, so that passengers who did not have experience of both trains were eliminated. It was also used to see, if people with more distant memories of the previous trains gave different results to those with recent memories. This it was hoped would give an indication of the effectiveness of time-series approaches.

Q.2 was a warm up question to establish rapport with the interviewee. It also served to get the passenger thinking about his decision to travel by train and to see whether ride quality was really an issue with the interviewee.

S.2 made the next questions specific to the Poole line, so that comments could be related to rolling stock that was familiar to the respondent.

Q.3 attempted to find words to describe the old and new stock. It also attempted to see, if the change in ride between the two stocks had been noticed. It thus went some way, to establishing how sensitive passengers are to changes in ride. By asking about quality in general, this question also looked at the perception of change. Ride may not have come out in this question, which would indicate its low priority in the change. If ride did not come out, it would have come out later in the more specific questions, so this lack of precision was not a problem. It was found that quality did not always include stock, so respondents were prompted with, "Thinking about the coaches themselves".

It should be noted that in all the questions the phrase, "If any" was used, so that respondents were not forced to produce issues by the wording of the question.

P.1 encouraged the respondent to exhaust his list of changes.

Q.4 was introduced to test whether changes in ride, if noticed, did make a difference to the respondents desire to travel by train. Q.4 thus gave an indication of the relative importance of the ride issue, for the design of the later surveys. There was an attempt to identify the direction of the effects, if any, as a result of the new train. Although commuters may say, "I have no choice than to use the train", they may still make off-peak trips where there is more choice. It was hoped that any deterrence effects of ride, would be identified here. With hindsight this question should have concentrated more specifically on ride. So many changes were noticed, that it was difficult to get respondents to give an indication of deterrence for each change.

P.2 was to remind passengers of the factors they believed had changed, as a result of the introduction of different trains.

Q.5 tried to identify what passengers associated with the word comfort. If perceptions were sufficiently narrow this word may be used later on.

Q.6 This represented an attempt to produce a phrase that meant, "Ride" to the general public. Roget's thesaurus was used in an attempt to find an acceptable phrase. During piloting the statement in this question was re-written a number of times to reduce the range of perceptions.

The first interviews used the phrase, "Involuntary movements of the passenger, caused by the train". "Movements of the passenger", concentrated on the passengers movements rather than the trains. This meant that respondents could not talk about the speed, or number of stops made by the train (as they did in the pilots) when asked about the movement of the train. "Involuntary" is used to eliminate movements that were not caused by the ride of the train, such as, walking to the buffet. The second part of the phrase states that, the movements are, "Caused by the train", as involuntary movements could be caused by other passengers on a crowded train. This therefore removes the effects of crowding from the statement.

It was decided that this movement phrase was a little complex and so a card was produced to display the phrase for the respondent to read as well. There were no problems with the phrase in the pilot, but the card was thought to be a useful safeguard. So that passengers would not think there was anything special about the movement phrase, a card was also shown for the comfort phrase.

This initial movement phrase was generally successful, though it was clear that passengers were having to think carefully about it. One or two people clearly found it too much effort and started talking about being pushed by other passengers etc. Some people just admitted they did not understand. As a result the phrase was changed twice during the interviews. Roughly half the respondents were tested using the original phrase and a quarter with each of the two replacement phrases. The second movement phrase left out, "Involuntary", as it was realised that if the movement was caused by the train, it must be involuntary. "Involuntary" was the word that appeared to be causing difficulty.

The second sentence was now found to give too little stress to forced movement. A greater proportion of respondents now began to talk about walking to the buffet, or failed to understand the question. The third and final phrase was most successful, "Being moved around by the motion of the train". Everyone understood this quickly and it appeared to cover all the elements of ride quality. Moreover people did not include anything other than ride quality with this phrase.

Q.7 This was an attempt to establish the language used by the passenger, to describe events on the train. It was also useful later on, when considering the deterrent effects of ride on passengers.

P.3 Was a prompt to generate more of the passengers own language and experiences associated with the movement of the train.

Q.8 tested two elements: firstly, if a ride improvement was not salient - was it noticed. It is suggested by Fishbein et al, that if such an attribute is not mentioned then it is not important. This question was intended to give some indication here. Secondly, whether a respondent can reliably remember the levels of comfort of the old stock. We may have got the reply, "I can't remember" or other non specific indications. This therefore gave us an indication of whether a respondent, would be able to gauge a change in ride between two trains over time. This possibility had to be considered if a time-series approach is used later.

7. ANALYSIS OF RESULTS:

The recording of each interview was transcribed and entered into a word-processor (for ease of reference), along with any notes made during the interview. Amenable data was then examined using a form of content analysis. Other information was grouped into categories and presented in a more general form.

The importance of content analysis was shown, by the difference between the interviewer's recollection of facts and those revealed through the analysis.

8. RESULTS OF THE INTERVIEWS:

8.1. Results - Language:

8.1.1. Ride:

As noted earlier, after three attempts a phrase was found that appeared to mean, "Ride" to all those interviewed. "Being moved around by the motion of the train", managed to extract a number of phrases to describe particular aspects of ride. Some alternative phrases were also developed by respondents. Listed below are the words/phrases used by each interviewee.

1). Streamlined, smooth-running, shaken-about, tipped-around. 2). Jolting. 3). Shake you all over the shop, ride. 4). Motion of the train, smoother. 5). Rocking-about, smoothness, smoother ride. 6). Smoother acceleration and braking, lurch-about. 7). Steadiness, smooth. 8). Smoother. 9). Getting jerked, smooth. 10). Smooth. 11). Jerk-about, better ride. 12). Smooth run. 13). Jerkiness, smoother. 13). Smoother ride, no jerks. 15). Floating. 16). More stable, jolted-around. 17). Train jolting, really smooth. 18). Smoother ride. 19). Heel, vertical acceleration. 20). Rocking about. 21). Heebie-geebies. 22). Jolted-about, smoother. 23). Smoother ride. 24). Jerked. 25). Shaken-around. 26). Jerked-about. 27). Swaying, lurching, smooth. 28). Comfortable ride, smoothness, jogged-around. 29). Smoothness. 30). Swaying from side-to-side. 31). Smooth journey. 32). Smooth journey. 33). Roll, ride, up and down motion.

It is clear from the phrases set out above that, "Smooth/ness" is the most common. 20 people mentioned this word out of the sample. 6 people mentioned, "Jerk/ed" and 4 mentioned, "Jolt/ing". As suspected, the technical term, "Ride quality" (used in places, by M.V.A) was not frequently used by the public. Smoothness appears to cover most facets of ride quality, while jerks and jolts technically only cover the rate of change of acceleration (m/s3). So, "Smoothness" may be of use in later work, while the other two phrases will be less useful.

Although, "Smoothness" seems a more elegant phrase than, "Being moved around by the motion of the train", the latter has now been tested. It may be that a, "Smooth journey" is also associated with a lack of problems on the trip. Perhaps the best solution would be the phrase, "Smoothness of ride": this has the precision of the technical term, while using the everyday language of passengers. This phrase will therefore be tested during the pilots of the next stage of the research.

Although trying to disaggregate a small sample is hazardous, it has been tried to examine relationships that may affect the results later. Doing this by class of travel and sex showed nothing. A significant variation with age was found. A null hypothesis can be rejected at a 95% level of confidence (Chi-squared 5.58, 1 d.f). We can therefore say, that people over 50 are less likely to use, "Smoothness" to describe ride quality.

8.1.2. Comfort:

It was thought that the public would perceive ride to be a large part of comfort, meaning that comfort could be a key phrase later on. But previous work has shown, that phrases like comfort and convenience cover a mass of variables and mean different things to each person. This research found similar problems with the phrase, "Comfort of the train". Some people narrowed the meaning down to, solely the seat they were sitting on. Others included: ride quality, buffet car, harassment by other passengers, ventilation, luggage, space, noise, reliability, parking, smoking, view, toilets, decor, cleanliness, ventilation, overcrowding, staff, and arrangement of seats. Generally ride quality was not seen as a major element of this phrase.

8.1.3. Old Train:

It was hoped that non-emotive ways would be found to refer to the old and new trains. The words, "Old" and, "New" were deliberately not used by the interviewer - unless the respondent had previously used them. The aim of this was to avoid the use of words that gave the impression of one form of stock being better than another. The words, "Previous" and, "Replacement" were successfully used during the interviews. But even these words may suggest the replacement is an improvement. The words volunteered, by the respondents, to describe the two types of train were in almost all cases, "Old" and, "New". Occasionally passengers referred to the trains by their official names, "REPS/TC's" and, "Wessies/Wessex Electrics".

Once it was established which train was which, respondents tended to refer to the characteristics of the two trains, using, "Now" and, "Then". This latter approach could easily cause confusion, especially if not all the previous stock has been replaced. Perhaps the most effective way of tackling this, is to use photographs of the old and new trains. The stock can then be referred to as, "This" and, "That" train. Obviously if photographs are to be used, care must be taken to ensure that lighting etc. is the same with each stock. In this stage of the research, a photograph was shown of the exterior of the previous train, plus the interior of a standard class coach (appendix two). Photographs were not shown of the replacement train, as all the interviews took place on these. It was therefore possible to say, "Think of the differences between this train and the one shown in the photograph".

8.2. Results - Sensitivity:

8.2.1. Change Noticed:

Ride quality was seen, by N.S.E, as one of the major changes with the introduction of the new stock. It is interesting to see to what extent the public agrees with this. The tables of changes below, show that ride was seen to have improved by most people. In fact sixteen people gave ride as one of the changes noticed with no prompting (seven of the sixteen suggested there had been a substantial change).

Of the remaining passengers, nine mentioned the change when asked about the, "Ride" phrase, and four noticed it only when asked specifically if it had changed with the stock. Thus twenty nine people, had to some degree noticed an improvement in ride. The others were unable to tell whether it had changed or not.

8.2.2. More Use Of Train:

Seventeen respondents who mentioned the improvement in ride were probed, to see if this improvement alone would make them use the train more. Only three people thought this was likely. In the general question (which asked about the features of British train travel, passengers were unhappy with) ride quality was well down the list. It thus seems that ride (at typical levels) is not a central factor in the decision to use the train.

Passengers were also asked if the introduction of the new rolling stock, would be likely to increase their use of the line. A large number of people commented, that they only used the train for work and so it made no difference. Such passengers were then asked if they would make more non-work trips as a result of the improvement in the service. Again very few suggested that they would, as most would still use the car for non-work trips. Generally it was agreed that the journey was more pleasurable with the new train, but that this would not cause them to make any extra trips. 11 people (33%) stated that they were more likely to use the train as a result of the improvements.

8.2.3. Past Experience:

Relating the sensitivity of the passenger, to the time since the they last experienced the previous stock, failed to show any statistically significant relationship. This is illustrated in the table below. It should be noted that no one suggested that they could not remember what the old trains were like.

TABLE 4.2: SALIENCE OF THE CHANGE IN RIDE.

	(LAST USED PREVIOUS STOCK) TODAY	IN LAST WEEK	LONGER
SENSITIVITY			
Volunteered	6	5	5
Semi volunteered	0	5	4
Asked	1	2	1
No change noticed	1	3	0

8.2.4. Ability To Isolate Attributes:

Statements made by passengers indicated that they were able to separate out the ride change from other changes with the stock. There seemed to be little expressed relationship between ride and noise - the, "Ride" phrases did not cause noise to be mentioned at all. Only two passengers stated that some of the ride improvement could be psychological, being related to an improved environment generally.

8.2.5. Salience:

An attempt was made to see, whether the salience of a ride change was indicative of its importance to the respondent. It was proposed that if ride was not freely given as one of the changes, the difference in ride was unimportant or not noticed. Eight respondents did not mention changes in ride before question eight and so were asked if they had noticed a difference. Of those asked: only one said it was a lot smoother and he had made earlier comments about the smoothness of the new trains (though these were not clear enough to count as a volunteered improvement). Of the remainder, three said there was a slight improvement, one considered only the jerk during starting to have reduced and three people failed to notice any change at all.

There is therefore an indication here, that if the change was not volunteered, ride was not an important (at the levels experienced by the passenger) or noticed issue. These findings may mean, that the salience of ride can be used later, to filter out respondents who find ride unimportant. However these findings are tentative and if this approach is to be used, more work may have to be done in this area.

8.3. Results - General:

Respondents produced many general comments about ride and other attributes during the depth interviews. These comments are now considered, both to give a broad view of the ride issue and an indication of its importance relative to other attributes.

For analysis respondents' comments are grouped into a series of categories - there may be some overlap between them. The frequency with which comments, about each attribute were made, are illustrated in the tables below. All comments were volunteered (by interviewees and others) during the interviews, or wind down period. Some other comments have been included that were made during off-the-record discussions. These latter statements are not recorded in the comment frequency tables. These results represent the issues salient to people during the interview period. The limitations of the sample, mean that care must be taken with interpretation.

The comments below are divided between general British Rail results (table 4.3) and those specifically related to the new trains on the Poole line (table 4.4). The general comments are heavily biased towards N.S.E. services, as this is where most of the passengers experience of British Rail is gathered. Results are presented in order of badness.

TABLE 4.3: COMMENT FREQUENCY (INTERVIEWS ONLY) GENERAL.
(ordered by net badness: bad - good comments)

	GOOD	BAD
Punctuality/cancellations		14
Overcrowding		11
Cleanliness		11
Passenger Information		5
Gen condition of stock		5
Buffet inc availability		5
Ventilation		4
Toilets		4
Journey time		4
Smoking accommodation	1 More	3 Less
Staff Knowledge		3
Luggage		3
Timetables		3
Staff lack of		2
Price		2
Smoothness		2
Noise		2
Personal security		2
Sunday services		2
Legroom		2
Staff untidy		1
Queues for tickets		1
Inconsiderate passengers		1
Telephone enquiries	1	1

It is clear from table 4.3 that three issues were considered particularly important by the sample: punctuality, overcrowding and cleanliness. Ride quality was not considered to be one of the most important issues, in fact it was equal fourteenth with only two adverse comments. These people saw the bumpy rough running of stock to be a problem.

TABLE 4.4: COMMENT FREQUENCY (INTERVIEWS ONLY) BOURNEMOUTH LINE.
(N.T. = New Train, ordered as above)

	IMPROVED	WORSE
N.T. legroom/seat size	1	8

N.T. changing at Bournemouth		5
N.T. exterior doors	2	6
N.T. layout/travelling backwards	4	7
N.T. luggage	1	3
N.T. claustrophobia/plasticy		2
N.T. seat design	3	4
N.T. reliability	2	3
N.T. cheap and nasty		1
N.T. buffet		1
N.T. cosiness		1
N.T. glamour	1	
N.T. litter facilities	1	
N.T. privacy	1	
N.T. crowded	1	
N.T. phone	2	
N.T. speed	2	
N.T. toilets	3	
N.T. newness/modern	3	
N.T. information	3	
N.T. space	4	
N.T. windows	4	
N.T. interior doors	4	
N.T. comfort generally	9	
N.T. decor (colour, carpets)	10	
N.T. cleanliness	10	
N.T. ventilation/draughts	13	
N.T. smoothness volunteered	16	
N.T. quietness		20

Considering the comparison between the old and new trains on the Poole line, ride was a much more significant issue. Ride was the second most positive change, generating sixteen favourable comments. Only the change in noise generated more (twenty) comments.

The smoother ride was noticed by most passengers. Jerks seemed to have been reduced and this was especially noted when starting or stopping. Some suggested that the improvement could be partly psychological, because of other improvements with the stock. A number of passengers made comments to the effect, that the change in ride was, "Not that crucial". Activities were often used to gauge the change in ride, for example, "You can now write on the tables" or, "It is easier to walk". This use of activity measures, meant that people could see the benefits of the smoother ride. Using such measures later in the research could be a problem as some passengers never drank or moved down the aisle. The ride on the old trains was characterised by remarks like, "Awful", "Getting thrown around" and, "People falling in to one another".

Some passengers noticed more swaying and rolling motions than previously. This slight long wave motion, it was said, could make one queasy.

8.4. Results - The Effects of Ride:

Drinking was seen as a problem by about half the sample. This was especially the case if the cup was too full. It was considered easier on the new trains.

Reading was considered a problem by about a third of the respondents. A number suggested that it was only a problem on the old trains.

Writing was also seen to be easier on the new train. This was commented on slightly more often than reading. Although writing was generally possible, a number of people stated that it would be difficult to read. These problems could make working on the train a problem.

Although walking did not come up as frequently as other difficulties associated with ride, this was partly because a number of interviewees did not have much experience of it. Even so, the problems of crashing into fixtures and other passengers were mentioned by about a quarter of the respondents. Some elderly passengers had stayed in their seats, because of the fear of the consequences of moving around to get cups of coffee etc.

A number of comments were made about the physical sensations associated with ride. These ranged from feeling sick (at the extreme) to annoyance and being tired/exhausted. Some people suggested that one got used to it as part of rail travel. Some passengers felt nervous because of the movements, especially when standing or walking.

Other comments were: That the movement of the train helped make you sleep, although one passenger complained that big jerks kept waking her up. Also, "Difficult to put on make-up", "Good for the circulation", "Gives a sensation of speed".

9. SUMMARY OF FINDINGS:

These preliminary interviews generated a series of findings, that will affect the design of the rest of the research.

Passengers were interviewed on new trains just introduced to the London-Poole line. Respondents took part in depth interviews that were based on a checklist of questions and prompts. These interviews were tape-recorded to allow checks for leading and other biases. The results were then transcribed and content analysed.

Thirty-three passengers were selected using a quota system. This meant that interviewees were roughly equally divided by sex and age. Interviews were done mainly on off-peak services, though there is some representation of peak users. Although this is a small probably biased sample, it is believed that the general information produced about passengers' views on ride issues is valid.

Photographs (appendix two) were used in an attempt to remind respondents of the previous trains. Carefully taken (representative) pictures proved the most effective way of referring to the previous trains in an unemotive way. All the verbal descriptions tested, either implied the old trains were inferior or confused the respondents.

The best phrase, developed during the interviews, to describe ride quality to the public was, "Being moved around by the motion of the train". The most common word used by people in response to this was, "Smoothness". It is therefore suggested that the term, "Smoothness of ride" should used (subject to piloting) in the remainder of the research. A, "Comfort" phrase was found to have considerably different meanings to each respondent.

When asked about the features of British Rail that made passengers unhappy, ride quality was well down the list. However ride was one of the most noticed improvements with the new stock. Most people (88%) saw ride quality as having improved, as a result of the introduction of the new trains. Women may be slightly less sensitive to the change. Ride quality, at current levels, does not seem to be a major issue in the decision whether to use rail. Though, a few people (33%) suggested the ride improvement may encourage them to make extra trips.

Approximately half the sample volunteered comments about the change in ride, with no prompting. There is an indication that if the change is not volunteered - ride is of less, but still of some, importance. This suggests, that in later work, all people should be asked to value a change in ride - even if they do not volunteer it.

Passengers' statements inferred that they were able to isolate ride from other attributes and surprisingly questions about ride did not cause noise to be mentioned. Although respondents generally considered themselves able to isolate ride, some mentioned that part of the improvement could be, "Psychological". A detailed analysis of ride ratings, between trains, may therefore display some contamination. This makes a further analysis of this issue necessary.

The most commonly elicited problems associated with ride were: about drinking, followed by writing, reading and walking. A few passengers were worried about the consequences of ride on such activities. As only about half the sample commented on the problems of drinking (the most frequently mentioned problem) a significant number of passengers appear not to have related ride to these activities. It therefore seems that trying to establish a universally applicable ride scale, based on such activities, would be difficult.

10. IMPLICATIONS FOR THE REST OF THE RESEARCH:

The information extracted during these interviews lays the foundations for future work. There is now a form of language with which to conduct further investigations and an effective unemotive way of referring to various types of stock. Passengers also seem to notice the change in engineering ride between the trains and seem to be able to isolate this ride effect.

These facts make it appear more likely that a stated preference approach - particularly one based on the differences in engineering ride between trains, is likely to succeed.

Chapter Five Design of Detailed Ride Investigation

1. INTRODUCTION:

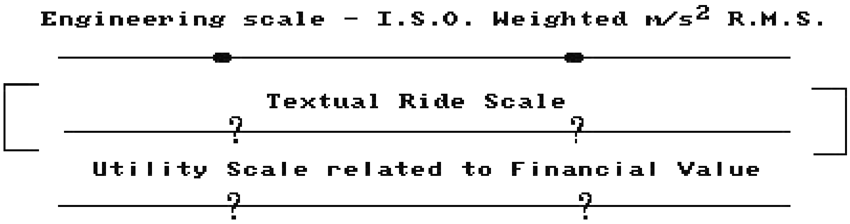
In this chapter the techniques that were considered likely to produce a value of ride (after the initial discussion in chapter three) are developed in detail. The findings of the depth interviews (chapter four) are applied in this context. The information needed to choose and implement a final selection of valuation techniques is also considered.

First general issues that will affect any ride valuation technique are discussed. Each approach is then investigated in turn. Although each possible approach is considered separately, there is a great deal of overlap between the techniques. Many issues considered in the earlier approaches, are relevant to those developed later.

2. RELATIONSHIPS OF SCALES:

As mentioned in chapter one, a major problem when attempting to place a value on a change in intangible attributes, is the difficulty of scaling.

FIGURE 5.1: RELATIONSHIPS OF SCALES.



This investigation attempts to place a financial value on changes in engineering ride, represented by the top scale. As stated earlier respondents are not familiar with the measurements of the engineering ride scale. We cannot therefore present them with changes on this scale to value.

There are two groups of ways of tackling this measurement problem. We can present respondents with directly experienced levels of engineering ride, for example the ride on the train yesterday. Or we can produce an alternative (textual) scale based on descriptions of ride.

This textual scale has to be related to the scale of engineering ride, so that we know what change respondents are valuing. There is no precise way linking these scales.

Once respondents have been presented with a change in engineering ride (in one of these forms) to value, they have to decide what difference this would make to their level of satisfaction. Individuals thus relate the change in engineering ride to a change in utility. There is further scope for error in this transformation.

From this discussion it is clear that the textual scale approach involves an extra transformation. This implies a greater potential for error. However this approach has a number of advantages over the directly experienced ones. For example the valuation sample do not need as much experience of real ride levels, this allows a more representative valuation sample.

When selecting valuation approaches we are concerned with the accuracy of these transitions. The issues that could affect this accuracy are discussed in detail in this chapter. Those issues about which little is known, will be investigated later in the research allowing a sensible choice of valuation approaches.

3. POTENTIAL ERRORS (WITH ALL TECHNIQUES) AND THEIR MINIMISATION:

To successfully establish a value of ride, all potential errors must be identified and then taken account of. Only stated preference approaches have passed the preliminary analysis and all are based on the model described above, thus some errors are common to all. Although it is possible that some of the errors may offset each other, it is unlikely that the extent of this offsetting will be known and it is therefore still important to minimise each error.

3.1. Sample Characteristics:

3.1.1. Sample Error:

Each individual is likely to have a different value for a given change in ride quality, we thus have to use the mean value to represent the population (all potential rail users). As we will only examine a part of the population, the sample's mean is likely to be different from the population's. The larger the variation of ride values, the greater the likelihood of sample error and therefore the less precise our results will be. This error is best contained by maximising the size of the sample.

3.1.2. Sample Bias:

It is likely that certain types of people and trips have consistently different values of ride. If the proportion of such groups in the sample is not representative of the population, results will be biased - regardless of sample size.

It is therefore important to note how ride values vary with personal and other characteristics. This would indicate any biases in the sample and allow ride values to be generated for a greater range of environments. It may be possible to produce some form of model based on these variations. For example, it is suspected that trip length and purpose will affect the values of ride held by respondents.

One of the most important influences on a sample's ride value is likely to be the regularity with which respondents have to stand. Standees are known to be more critical of ride than seated passengers. It is therefore likely that passengers with greater experience of standing will have higher values of ride. This issue is of increasing importance because of greater overcrowding, as a result of extra demand not matched by stock replacement.

A seating verses ride trade-off could become an important issue in this thesis; the more seats and the less the need to move around, the poorer ride can be. Further investigation into this area may be worthwhile. Previous work has only compared respondents' perceptions while they were experiencing the ride. A travel decision is likely to be made after the event and on the basis of more than one experience. It would therefore be useful to compare individuals' perceptions of ride after the event, with the number of times they have stood.

Closely related to the sampling problem is that of non/infrequent-users, it is expected that the latter may be differently critical of ride quality than more frequent users. Non/infrequent users may thus have to be taken account of in the research design. This could mean that the valuation exercises have to be done as a household survey. One obvious technique for securing a representative sample of such users would be to construct a representative panel of respondents. These people would be sent on a series of journeys - ensuring that they had sufficient ride experience. But the restrictions of a Ph.D. and the need to avoid hyper-sensitising respondents makes this exercise impractical.

It is clear that sample bias could have an important adverse effect on the results. This error is best tackled by ensuring that the structure of the sample reflects the population. If this is not possible the size and direction of the likely effects should be determined, so that results may be compensated.

3.2. Environmental Effects:

Related to sample error is that caused by situation. It is suggested in previous research (chapter three) that decisions can only be accurately replicated in the decision environment. If this is the case with travel decisions, we must find where the decision to travel by train was made. We must also ensure that any questions about the travel decision are asked in the appropriate environment. Meeting these requirements would make the design of any valuation exercises more complex. For example all valuation approaches may have to be implemented using home based interviews.

An investigation into environmental effects therefore seems justified. We must establish whether there is any difference between the points elicited in the decision environment and those gained on-train.

3.3. Interference Effects:

Earlier work has shown that respondents may have difficulty isolating and rating a single intangible attribute (like ride) because of interference effects. There are three related forms of effect (discussed in chapter three). If we are to value a change in ride quality an attempt must be made to contain, "Halo", "Newness" and, "Contamination" effects.

Interference effects occur when the general impression towards an object colours an assessment of a particular attribute. Basically, ratings are contaminated (Guilford (1954)). This makes it difficult for a respondent to isolate one attribute and make a reliable assessment of it. It has been suggested that ride quality and noise are connected in this way. So two trains with the same engineering ride, but different levels of sound insulation, could be assessed as having different levels of ride quality by passengers.

Guilford, when reviewing rating scales, states that such effects are worse where:

a). The trait is not easily observable.

b). The trait is not normally singled out, or discussed.

c). The trait is not clearly defined.

d). The trait involves reactions with other people.

e). The trait is of moral importance.

The first two categories clearly relate to ride quality, while the third could. Interference effects could therefore have a significant effect on the validity of the results and thus on the research design.

Although there are considered to be a number of different types of interference effect, the distinction between them is not entirely clear. For example a new train would achieve higher ratings than an identical old one. This difference in ratings could be considered the result of a Newness effect. But the new train would probably be cleaner, perhaps resulting in a Halo or Contamination effect.

The existence of interference effects means that it is important that typical days (with no abnormal events) are selected for surveys. All stated preference surveys will be biased by peoples disposition at the time (a bad day at the office) but this is a random bias, one caused by a a late train etc. is not and could reduce the accuracy of the results.

3.4. Other Effects:

It is important to consider whether the relationship between changes in engineering ride and financial values (via utility) is linear. If this is not the case, a given change in engineering ride may have a different financial value at different points on the engineering ride scale.

It should be stressed that all stated preference approaches are dealing with hypothetical decisions and not observed behaviour. So any results could be falsified or just incorrect (Moser and Kalton (1971)). As respondents' tasks could be related to policy changes, it is possible that responses may be exaggerated in an attempt to obtain improvements. To minimise such difficulties it should be ensured that questions are relevant to the respondent, non-leading and unambiguous. Hypothetical decisions should be as simple as possible, whilst being realistic. Valuation exercises should not be too, "Obvious" - reducing respondents' opportunities to manipulate results.

4. TEXTUAL SCALE APPROACH - ON-TRAIN AND HOUSEHOLD:

4.1. Introduction:

This approach involves a unique problem. As with this technique respondents have to perceive levels of engineering ride from locations on a textual scale; with the other techniques, levels of engineering ride are directly experienced. The difficulty of this translation is illustrated in M.V.A's report (May 1986): passengers were asked to select the most appropriate textual ride description for the train they were on - the chosen descriptions were widely spread.

With such an approach it is important to note, that we are not assuming that respondents say, "This location on the textual scale means 2 m/s2 R.M.S. to me". But that individuals will visualise certain sensations from a location on the textual scale, which are a proxy for an engineering level of ride. To be able to use any value of ride produced by the research, we must be able to find the engineering measures for which the sensations are a proxy.

As mentioned in section 2, interpersonal variations mean that it is likely, that each respondent perceives the relationship between engineering ride and the textual scale differently. For a valuation of ride to be successful, we must be able to associate a single engineering ride measure with each location on the textual scale (which is not easy) and recognise this is an average.

4.2. Development of the Textual Ride Scale:

Initially the most promising way to generate a textual ride scale appeared to be to develop a universal textual scale based on activities (for example: sickness, inability to read, write, walk). Descriptions could be given to respondents that were directly related to engineering ride measures. However the depth interviews have shown that this approach would be unlikely to succeed. Although one can say that certain activities would be more difficult than others (at a given level of engineering ride) respondents differed in their ranking of the difficulty of activities and a number had no experience of some activities on-train.

Many activities only relate to certain aspects of ride. For example drinking problems are generally associated with peak movements or jerks; sickness is often associated with long wave up and down motions. This means that different activity measures are not directly comparable as they are measuring different elements of ride quality.

It is thus clear that the only way to proceed with the hypothetical approach is to use a textual scale that is not derived from engineering ride measures - similar to that used by M.V.A (May 1986). An example of such a scale would be: very smooth ride, smooth ride, acceptable ride, rough ride, very rough ride. Although M.V.A's scale was not quite this imprecise, the difficulty of relating their findings to engineering ride measures has proved a significant obstacle in the implementation of their results. In M.V.A's case the levels of engineering ride which corresponded to the textual descriptions were arbitrarily chosen by British Rail. Thus despite the care taken by M.V.A. to produce an effective stated preference design, the small value of ride and the imprecision of the relationship between textual and engineering measures makes their results almost meaningless.

As M.V.A's textual descriptions are not related to specific trains, their results do not contain any interference effects. But this research considers it more important to have ride values that can be applied, than to totally eliminate interference effects. Even so with textual locations developed from the same stock running over different qualities of track, any interference effects should be controlled. The only difference between textual ride locations would therefore be due to ride quality (section 4.7.3).

A textual ride scale (that can be effectively related to engineering ride measures) will thus have to be developed if the hypothetical approach is to progress.

4.3. Optimising a Textual Ride Scale:

4.3.1. General Issues:

We must ensure that the textual ride scale is constructed in a way that minimises error, so that any interpersonal variations in ratings are caused solely by differences in tastes. Oppenheim (1966) states that, "The use of ratings invites the gravest dangers and possible errors, and in un-tutored hands the procedure is useless". Meister (1985) agrees that rating scales are vulnerable to biases and errors. He also states that graphic rating scales are harder to score and their results may imply a degree of precision and accuracy which is unwarranted.

However Meister (1985) concludes by saying that rating scales, "Can reflect both the direction and degree of opinion/attitude. The results are amenable to analysis by conventional statistical tests. Graphic rating scales permit as fine discrimination as the respondent is capable of making. They usually take less time to answer than other types of items, can be applied to almost anything, and are generally more reliable than two-way multiple choice items". The errors and variations likely from using a textual ride scale are considered below, in the order in which they would be generated by a respondent.

First the respondent has to establish that he is required to give his personal rating to certain aspects of a rail journey. It is therefore important that the instructions and descriptions along the scale are unambiguous and easily understood. Clear and concise everyday words must be used in place of the technical jargon associated with this subject. The information collected in the depth interviews and piloting of the proposed designs should ensure these requirements are satisfied.

Second the respondent has to establish what he is required to assess. It is important that each person considers ride quality alone while carrying out the scaling exercise - so that the textual measures developed are valid. It follows from this that each respondent should see the task in the same way; thus care has to be taken to ensure that the instructions and descriptions along the textual scale are interpreted in the same way by a variety of respondents. Any differences in locations recorded on the scale must be (as far as possible) purely a reflection of differences in engineering ride levels.

There is bound to be some interference from other attributes (section 3.3) but it is believed that the depth interviews have provided sufficient information to create a scale that minimises such problems. Other interference is controlled by randomisation. For example a ratee in a bad mood is assumed to be balanced out by other ratees, who are in a better mood. These problems are discussed in more detail later.

Finally the respondent has to consider where to place his ratings on the textual scale. An important consideration here is the way in which the scale is anchored. Anchors are the labels attached to rating scales at specified points, enabling respondents to place assessments at a location on the scale. For example the scale illustrated later has two anchor points: "Very rough ride - Drinks spill frequently" and, "Don't notice train is moving". Unfortunately such anchors are likely to be associated with different levels of engineering ride by each respondent. Some people would see, "Very rough ride" occurring at lower levels of engineering ride than other people. The reliability of the scale is closely related to this - it is important that if the same assessment was done a number of times by a respondent, the same location would result.

4.3.2. Scale Construction:

Rating scales can be divided into two groups. There are continuous graphic scales, where respondents mark a position along a line and discrete scales where respondents mark one of a limited number of options. The issues involved in constructing such scales, in the context of this research, are now discussed.

Each person's rating of ride quality is likely to be affected by their experience, which will modify their beliefs of what is possible and acceptable. This suggests that individuals with different experiences of travel may rate levels of engineering ride differently.

It is likely that the acceptable level of ride quality increases over time. This point is illustrated by a study carried out in the late 1960's (Research Projects (1968)) when 4-VEP rolling stock was introduced. The respondents in 1968 considered this, then new, stock to be characterised by a smooth ride. However in later research (Steer, Davies and Gleave (1983)) respondents gave similar stock a poor ride quality rating. This is despite likely improvements in track and continued maintenance of the old stock. Thus a mean textual rating of 4.2 for a particular engineering ride level today, may fall to 3.2 for the same experience in ten years time. These changes must be considered if such ratings are to be used over time.

It is important that the experience gained from the previous use of rating scales, is used to enhance the design of the textual ride scale. Guilford (1954) lists the errors usually associated with rating scales:

a). Error of Leniency: When producing assessments of individuals, respondents rate people they know higher than they should. Some respondents are aware of this and over-compensate, giving lower ratings than they should. Leniency errors are greater if results are not confidential. Leniency increases with length of acquaintance and men are more lenient.

b). Error of Central Tendency: Raters hesitate to give extreme judgements, this is very common where the subject is not known that well. The scale can be developed to spread ratings further out.

c). Interference Effects: Described earlier.

d). Logical Error: This occurs where respondents feel that attributes are logically related, so if one attribute is high another, "Should be".

e). Contrast Error: When rating individuals there is a tendency for respondents to emphasise the differences from their character, so an orderly ratee would consider others to be disorderly. However an opposite bias is also possible where a respondent expects others to be like himself.

f). Proximity Error: Caused by the nearness of rating two traits in time, or space. One rating contaminates the other (otherwise known as order effects).

g). Un-trained Respondents: Lack of rating experience can reduce levels of accuracy. Respondents can be trained to reduce this problem. I.Q. is positively related to the quality of ratings.

Other findings noted by Meister (1985) and Guilford (1954) include: The more time available the more accurate the ratings. The sureness of a ratee is an indication of their accuracy. Knowledge of the purpose of the exercise can affect ratings and different ratees use different criteria in rating the same trait.

Considering graphic scales, Meister (1985) states that, "Variations in scale format (horizontal verses vertical etc) have produced statistically significant effects, but practically speaking the effects are unimportant".

However Guilford (1954) argues that vertical graphic scales are the most effective. With vertical scales there is more room for anchoring descriptions and the descriptions are not spread along the scale. Guilford states that, the line on the scale should be at least five inches long, but not much longer. Longer lines can lead to clustering, while short lines do not allow enough discrimination. The line should have no breaks or divisions, as a continuous line suggests continuity of the variable (scoring can then be decided by the interviewer).

Guilford states that with a vertical scale the good end should be at the top, which is what the respondent expects (like a thermometer). With a horizontal scale, the good end should be on the right. If a series of scales are to be used, it is sometimes suggested that the good and bad ends are alternated to reduce errors. But this has not been proved and it confuses respondents. To avoid subjectivity in coding, a stencil should be used for scoring positions on the scale.

Anchors have been found to improve scale reliability. But it can be difficult to develop anchors for the central part of scales. Behavioural anchors (developed by a group of judges) can further increase accuracy; but overall Meister (1985) concludes that the extra effort needed to produce such anchors, is not worth the minimal increase in accuracy. Guilford (1954) states that anchors should be:

a). Clear: Short and unambiguous. It is important that a respondent's interest in maintained. As more interested ratees give more valid ratings.

b). Relevant: Consistent with the trait name and conveying no implication of other traits.

c). Precise: They should apply to a point, or short range on the continuum.

d). Varied: Language should be varied at different points on the scale.

e). Objective: They should be not use moral terminology, implications of good or bad should be avoided. Descriptive scales suffer less leniency than evaluative scales.

f). Unique: Anchors should be unique to that trait.

Guilford (1954) makes some further points. There is a tendency with graphic scales, for assessments to cluster around the anchors. Anchors do not need to be evenly spaced and are sometimes spread more in the middle to counteract the error of central tendency. For this reason anchors must not be too extreme.

4.4. The Chosen Scale:

With these issues in mind we now have to decide on the most effective form of rating scale. A continuous graphic textual ride scale rather than a discrete one, used conventionally and by M.V.A, appears the most promising.

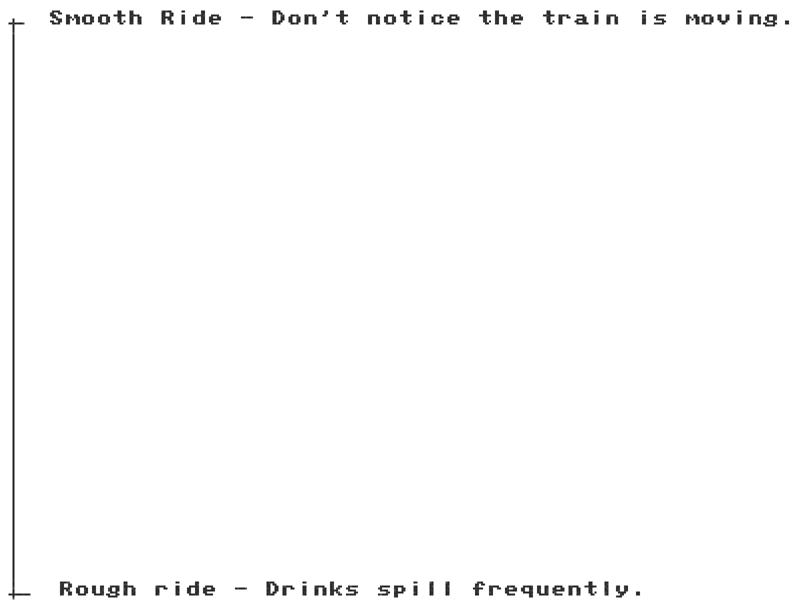
A continuous graphic scale infers that the variable is continuous; respondents are able to place their assessments on the scale spatially - making a linear interpretation more valid than with discrete scale responses.

A continuous scale allows the respondent a greater degree of discrimination between ratings, this is especially important when considering attributes that may make only a small difference to satisfaction. Such a scale avoids the problem of how many categories to select. The difficulty of finding anchors for a discrete scale increases with precision, this is not the case with a continuous graphic scale. A continuous scale is not therefore as dependent on an individual's perceptions of language, as a discrete scale with many anchors. From this it seems that in those cases where a variable is to treated as continuous, a continuous scale provides the most effective form of rating.

It has been found that graphic scales can correlate as strongly as discrete scales with objective measures. Graphic scales can exhibit slightly more variation in measurement and they tend to be more difficult to administer and comprehend by the respondent (Friedman and Friedman (July 1986)). However in this case it is considered that the benefits of the graphic scales outweigh these disadvantages. The likely scale is shown below.

FIGURE 5.2: SUGGESTED RIDE SCALE.

How smooth is the ride on this train?
(Mark a position on the scale with an X)



Every respondent would place each requested level of engineering ride experienced, at a location on the textual scale. If the textual scale is assumed to be linear a sample's mean location can be derived for each level of engineering ride desired. For example, the mean for a Sprinter may be located 6.3 units along the scale and a Pacer 4.2. These mean textual locations are directly related to engineering ride measures - which can be measured using a Macmeter.

The mean locations on the textual scale can be presented to respondents later in the same format, causing them to perceive engineering ride levels for valuation. This bi-directional process is only possible with a continuous scale of this type. It has to be assumed here that respondents are able to work in both directions with the textual scale: that is, they can perceive levels of engineering ride from locations on the scale as well as recording levels of engineering ride on it. This assumption is unavoidable if a ride valuation is to be extracted using the textual scale approach.

Two anchors will be used, one at each end of the textual scale. Although many rating scales use three or more anchors, there has always been difficulty in establishing anchors for the central part of scales. If a respondent does not perceive a central anchor to be in the correct location, this could make the rating task more difficult as the scale may appear to be kinked.

Any assumptions of linearity therefore become much more tenuous, when central anchors are used. Although having too few anchors on a rating scale can make rating less precise, it is considered that the problems caused by central anchors are greater than those caused by having too few anchors on the scale. Finally for the analysis to be successful, the difference between the anchors must be sufficient to generate a spread of ratings along the scale; it must also be ensured that everyones' assessments fall between the anchors.

The most effective approach to anchor design, is to use the information elicited from individuals in the depth interviews. This is similar to an approach described by Meister (1985) known as, Behaviourally Anchored Rating Scales (B.A.R.S). These differ from standard graphic scales because of the addition of anchoring illustrations which are, "Concrete and specific". B.A.R.S. are designed to standardise the raters observations and screen out idiosyncrasies.

"Smooth ride" was the most commonly used term in the depth interviews; it therefore seems sensible that the both anchors should be derived from this term. But, "Smooth ride" is very subjective and so it is proposed that the description, "Smooth ride - Don't notice the train is moving" will be initially selected as the first anchor. Although this latter phrase is still open to different interpretations, the range of perceptions will probably be substantially reduced.

Similarly at the other end of the scale, some universally accepted characteristic has to be found to reduce the range of perceptions and increase reliability. The most commonly mentioned characteristic of a very rough ride, during the depth interviews, was drinking. The initial anchor for the rough end of the scale will therefore be, "Rough Ride - Drinks Spill Frequently".

4.5. Application of the Textual Ride Scale:

4.5.1. Importance Of Minimising Standard Errors:

As stated earlier each person is likely to perceive the relationship between the textual scale and engineering ride differently. It is therefore likely that any sample of individuals will generate mean locations, on the textual scale, that are different to those that would be generated by the population (all potential rail users).

It is important to minimise the possibility (standard error) of a sample's mean textual locations not matching those of the population. If this is done, when the sample's mean textual locations are used to establish a value for ride, the valuation sample are more likely to perceive the correct engineering ride levels. If the correct engineering levels are not perceived, the ride valuation extracted may not be appropriate.

If the relationship between engineering ride and the textual scale is very weak, this approach could be invalidated. Other proposed approaches directly compare engineering levels of ride (experienced by the individual on various journeys) and so would be immune from this problem. However alternative approaches are associated with other difficulties.

4.5.2. Size And Representativeness Of Sample:

The obvious way to ensure that a sample's mean scores on the textual scale match those of the population, is to use the largest practical sample. It is also wise to ensure that any sample is representative of the population. But the relationship between the textual scale and engineering ride would inevitably be established with (an) unrepresentative sample/s. Experienced travellers would dominate as respondents would be required to relate either previous, or the current trip/s, to the textual ride scale.

Establishing mean textual locations with (an) unrepresentative sample/s may not cause problems, if individual's ride perceptions are consistently better or worse than the population mean. If perceptions are consistent the valuation sample will always be considering similar sized changes on the textual scale, though at different points.

However if the value of a change in ride is not linear, such an unrepresentative textual scale could generate inaccurate ride values (section 3.4). It is therefore important to consider, whether peoples perceptions of engineering ride are consistently above or below the population mean and whether values of ride are linear.

4.5.3. Controlling For Changes Between Scale Construction And Application:

For the valuation exercise to be successful, it may be important that the textual scale is related to engineering ride in the same environment as the ride valuation exercise. If this is not done, the valuation sample may not perceive the original engineering levels from locations on the textual scale.

As suggested in the previous section, it may be that such factors work consistently on all textual locations. So the differences between textual locations may be similar wherever the scale was developed. The accuracy of the ride valuation may therefore not be affected, unless the valuation is significantly non-linear.

Even so it would be wise to check the extent to which assessments of ride change with the environment. Generally we must try to keep all environmental and sample factors constant between the development of the mean textual ride locations and the valuation exercise.

4.6. Number of Textual Ride Levels to be Valued:

The number of textual ride levels produced is limited by what can be achieved in the time available and the ability of the sample/s to accurately distinguish between engineering levels. If the same people are expected to consider all the selected engineering ride levels, the exercise is also limited by respondents' experience.

Two engineering levels have to be considered to produce a valuation - three or more if the linearity of ride values is to be investigated. The most likely number of levels to consider would be three or four (as used by M.V.A (May 1986)). As this research is directed at rural services, it may be appropriate to consider a number of Provincial trains, for example: Sprinters, Pacers and old Diesel Multiple Units. Further information is required before a final decision on this point can be made.

4.7. Alternative Research Designs:

4.7.1. General:

To relate locations on the textual ride scale to engineering measures, we need respondents who have experienced specified levels of engineering ride. There are two groups of ways of selecting engineering ride experiences to establish this relationship. We can use repeated measures where one sample assesses all engineering levels of ride, or independent measures where each level of engineering ride is assessed by a different sample.

REPEATED/DEPENDENT MEASURES:

a). Getting each individual to assess ride over various sections of one journey.
b). Each individual assesses several alternative journeys, some or all from memory.

INDEPENDENT MEASURES:

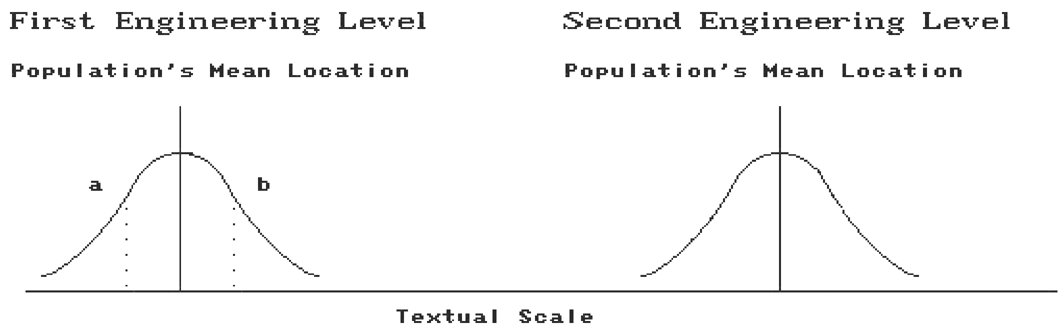
c). Each individual assesses only one journey or section of track.

Each of these designs are now investigated in detail.

4.7.2. Repeated Verses Independent Measures:

The way in which the textual positions are generated can effect the scale's accuracy. It is therefore important to determine whether independent or repeated measures are most likely to generate representative positions on the textual scale.

FIGURE 5.3: INDEPENDENT MEASURES.



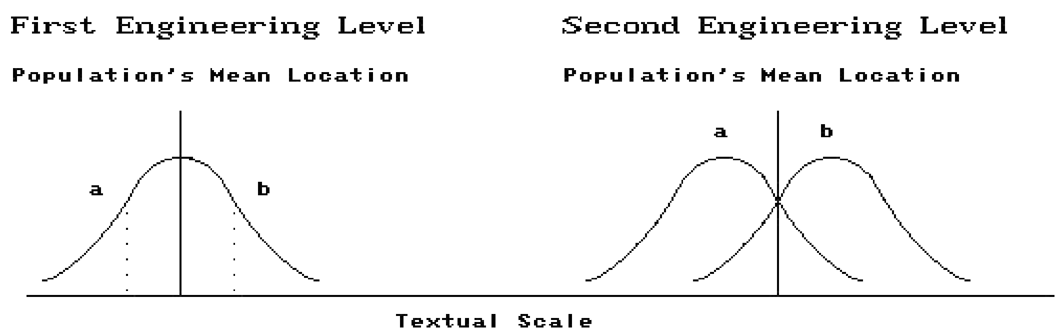
Figures 5.3. and 5.4. show the effects of sample type on the accuracy of the estimated textual locations. For both types of sample the first assessment is free to vary, so the distributions of possible mean locations are identical and both centered around the population's mean.

With an independent sample the second distribution of mean locations is again free to vary and is thus also centered around the population's mean (figure 5.3). However the distribution of mean locations for the second dependent location will be offset from the population's mean (figure 5.4) by the same amount as the first location was above/below the population's mean. This occurs because the personal characteristics that influence the severity of the first assessment are still present during the second assessment. Thus individuals who rate the first engineering ride level above the norm (b), are most likely to rate the second level above the norm (for example, Miller (1984)).

The extent to which individuals' ratings are consistent will determine the standard error of the distribution a/b. We do not know how consistent respondents will be in this context. We therefore cannot estimate the relative sizes of the standard errors for the dependent and independent distributions of second mean locations. This means that it is not possible to state which approach is likely to yield more accurate textual locations.

As we are valuing the difference in ride on the textual scale, it is important that the difference between the textual locations is perceived correctly. Repeated measures are more likely to do this, as any over/under assessment of the first engineering ride level will result in a similar over/under assessment of the next level. If an independent first assessment is over/under the population's mean this compensation not occur - generating an inappropriate difference on the textual ride scale.

FIGURE 5.4: DEPENDENT MEASURES.



From this discussion it is clear that repeated measures are just as likely to match the population's mean locations on the textual scale. They are also more likely to generate a representative difference on the textual scale for any pair of engineering ride levels. A textual scale generated by repeated measures, is therefore more likely to produce a value of ride appropriate to the change in engineering ride under consideration. But it should be remembered that there are other issues affecting the choice of sampling approach. These are now discussed during the development of the various experimental designs.

4.7.3. Dependent (One Journey):

This approach means finding a group of travellers on a route with a number of sections characterised by different levels of engineering ride. With this method each respondent is able to give a series of ride assessments, soon after the experience.

This approach is likely to suffer from order effects; these occur where a respondent is exposed to a series of experiences, each of which affects the perception of the following one. It may be that as the number of ride assessments, produced by a respondent, increases the more fatigued he becomes. Fatigue may bias the results either way. Order effects will probably be more severe, if assessments are made during the trip, rather than at its end.

If order effects are symmetrical, the problem can be reduced by testing one half of the respondents in the opposite order to the others (Miller (1984)). Such a compensation is unlikely to be effective in this case as it would involve interviewing passengers on trains travelling in both directions - engineering ride is likely to change with direction.

It is likely that respondents will have difficulty in distinguishing ride sections from each other. Even if respondents are asked immediately at the end of the section concerned, they will still be using their memories of that section to make an assessment. Memories of earlier sections of route will interfere with the current assessment. As all the experiences occur at a similar time and place, the memories will be less easy to distinguish than if the experiences occurred separately (Baddeley (1982)). This problem may be slightly worse if (to avoid hyper-sensitivity) respondents are asked about individual sections of route only after the whole trip. Apart from this major problem, recall should be more accurate with this design as the experiences are more recent.

The problem of hyper-sensitivity has been touched on earlier. If respondents are aware during the trip that they are assessing ride, it is suspected that they will be more critical. Hyper-sensitivity can be removed if all the assessments are done after the train has passed the sections concerned. Also, producing assessments of ride during each section requires a much greater effort from the respondent. They will be involved in the exercise for a much greater part of their trip. Unless each respondent is closely supervised (which would severely restrict sample size) we cannot be sure that each ride assessment has been done at the end of relevant section. Overall this suggests that assessments should be done after all relevant sections have been passed.

It may be difficult to get a sample of people that are riding all the sections of the route we are interested in. A line would thus have to be found with substantial through traffic; this may eliminate some routes with other beneficial characteristics. It may also be difficult to find a service that covers a number of sections of line, each with different engineering ride levels.

On the positive side engineering ride is the only attribute that will have changed between the assessments, which makes the isolation of ride relatively simple and thus more accurate. With other approaches, not all of the difference between textual ride locations may be due to different levels of engineering ride, some may be caused by changes in other stock attributes.

4.7.4. Dependent (Many Journeys):

Ideally this approach would be implemented with either all the individuals' assessments based on previous trips, or none at all. If only some assessments were based on previous trips an extra (unwanted) variation would be introduced into the experiment. Research has shown that our memory of events reduces rapidly, though the rate of memory loss slows considerably: 40% within 20 minutes, 55% in an hour, 65% in 24 hours and 70% in 31 days (Baddeley (1982)).

Basing all assessments on current trips has been discounted (apart from the previous method - section 4.7.3) as we do not have the resources to get the same people to experience various levels of engineering ride. As some assessments will have to be based on previous (different) ride experiences, we must use an experienced (biased - see section 4.5.2) sample of respondents.

Experienced travellers would not be efficiently picked up with a household survey. Contact thus has to be made on stations, or trains. Interviewing passengers on platforms is difficult - time is limited as people have deadlines to meet. Further, only people with loose schedules would be successfully captured, which biases the sample in an unknown way. It would be possible to give out self-completion questionnaires, but there would be no control over the environment in which they were completed (which may affect the scaling of the results) and the response rate would be low as questionnaires would have to be returned by post (the returns may also be unrepresentative). This leaves two possibilities: collecting addresses on the platform and conducting interviews at home, or interviewing respondents on trains.

It may not be easy to get respondents to agree to be interviewed at home, it would also be difficult and time consuming to reach those who did agree. Home interviews also suffer as it is suggested (by Baddeley (1982)) that an environment appropriate to the topic makes recall easier. It is therefore likely to be more difficult for respondents to recall ride quality in a totally non-rail environment. If this dependent approach is to be used, the interviews will therefore take place on-train.

If all ride assessments were based on previous trips, the current train would not be used in the generation of the textual ride descriptions. Two or three other trips would be recalled for the interviews. This may confuse respondents. The best approach would probably be to ask for a rating of a section of the current trip immediately after it had occurred, as well as a series of assessments of previous trips.

With all assessments based on previous trips or an earlier part of the current trip, no hyper-sensitivity will be present with this approach. Respondents will have experienced ride levels in the past, before they were aware that they were important to anyone.

The recollection of previous trips is likely to suffer from interference, both from the current trip and memories of other previous trips (section 4.7.3). This means that respondents may not be able to relate memories of engineering ride to specific stock and routes. Generally, "Designers of surveys and questionnaires may be tempted to rely on our memories to a greater extent than is wise... expecting us to respond with a degree of detail that is totally unrealistic" (Baddeley (1982)).

To be able to aggregate any results we want all respondents to recall the same set of engineering ride experiences. It may be hard to find a group of respondents who have the required experience. Even if every respondent has ridden the same trains on the same routes, there may be some variation in the engineering ride quality experienced between individuals. Some passengers will have had worse ride experiences than others: for example, the track/suspension may have been particularly bad, or the train may have been running faster to recover time etc. As long as such variations are random across the routes/stock there should be no problem as any effects would balance out over a large sample.

There may be some difficulty in establishing the appropriate measures of engineering ride for the locations on the textual scale. Previous trips may have been made on stock that is no longer available on a given route (or at all) to measure.

The quality of track will vary along routes. If respondents are to assess the ride over whole routes, we cannot be sure that they have experience of the whole route. As patronage will also vary along a route, asking a group of respondents to give an overall assessment of ride may produce a biased result.

Asking respondents to recall ride on certain sections of routes is unlikely to be successful. We are not sure whether respondents could isolate ride on sections of the current trip. Recall error would make it even more difficult to isolate such similar experiences on previous trips (Baddeley (1982)).

We could ask respondents to compare the ride of different trains on specified routes. But this would require a line where stock is regularly mixed. This would also cause ride ratings to be interfered with by changes in other attributes (see section 3.3).

4.7.5. Independent:

With this approach each respondent assesses just one engineering ride level. Textual ride locations are produced for a number of engineering ride levels: by interviewing different sets of respondents for each engineering level.

The time required for an interview and finding respondents experienced in appropriate ride levels, again suggests an on-train approach. Respondents would be asked about the current trip rather than a previous one: asking about previous trips may confuse the respondent, reducing accuracy. Further, previous trips will not be recalled as accurately as the current one: the former will be a more distant memory and may be interfered with by current events. Asking about previous trips also makes sampling more difficult (and more biased) as we are increasing the level of experience needed by respondents to be able to answer the questions.

As we are asking about the current trip, we are certain that every respondent has adequate experience and knows what we are referring to. With memory based approaches we cannot be sure the respondent is thinking of the same situation that we are. The above means that this approach probably produces the most accurate recall of ride quality.

Ideally an assessment of the current trip would be made at its end to avoid hyper-sensitivity. This may be difficult as the respondent will be leaving the train and will not have time to complete a questionnaire. If questionnaires were issued at the end of the trip, they would have to be returned by mail later. But with a mailed return there is no control over when and where the assessment is made, or how much attention is paid to it. Further, it may not even be filled in by the respondent. A low response rate typical of such approaches may also cause difficulties. Such risks of off-train completion seem likely to cause more problems than hyper-sensitivity.

One possible way around this difficulty would be to ask each respondent about one section of their current trip. People could then be asked for their opinion of the section before they left the train. This approach does however suffer if people are unable to isolate the ride of one section from earlier part of their trip.

If this approach is used, the assessments will therefore be done on-train for the current trip. However some portion of the questionnaire may be completed later, allowing responses under various conditions to be considered.

With an independent approach respondents do not make multiple assessments. This means that more interviews have to be done in order to achieve the same number of engineering ride assessments. However with single assessments more engineering ride levels can be considered, as we are not restricted by the experience of individuals. Also as each respondent is only concentrating on one experience, recall and accuracy may be improved. The lower emphasis on experienced respondents means that a broader (and more realistic) sample of people can be used.

One problem with this approach is that it does not allow an investigation of interpersonal variation in the perception of changes in engineering ride.

The questionnaire is likely to more straightforward with this approach as we are only dealing with one train. This also means the whole process will be quicker and less confusing as only one trip is considered.

4.7.6. Conclusions On Experimental Design:

With both dependent designs, as each person will have produced two comparable locations from memory, interpersonal comparisons of the perceived changes in ride quality can be made. As we wish to know whether people are consistent in their perceptions of engineering ride (see section 4.5.2) it is important that at least some textual locations are derived using a dependent approach. The dependent assessments also benefit over the independent ones as they require fewer interviews and a textual scale based on such assessments is believed to be more likely to produce an accurate value of ride (see section 4.7.2).

The best of the two dependent approaches is probably that based on different sections of one trip (section 4.7.3). Although this suffers from order effects, it is considered that the reduced recall error, lack of interference from changes in other factors, lower need for experienced respondents and lack of confusion as to which engineering ride levels are being assessed, makes this the more promising approach.

The difference in engineering ride between track sections may be smaller than that between trains or suspensions. This may mean that the preferred dependent approach (section 4.7.3) is unable to provide locations on the textual scale that are sufficiently discrete. If this proves to be the case, one of the dependent assessments can be treated as an independent assessment and further interviews can be conducted to allow locations to be developed on another type of train.

5. ACTUAL CHANGE IN RIDE - STOCK CHANGE - STATED PREFERENCE - TIME-SERIES:

5.1. Basic Approach:

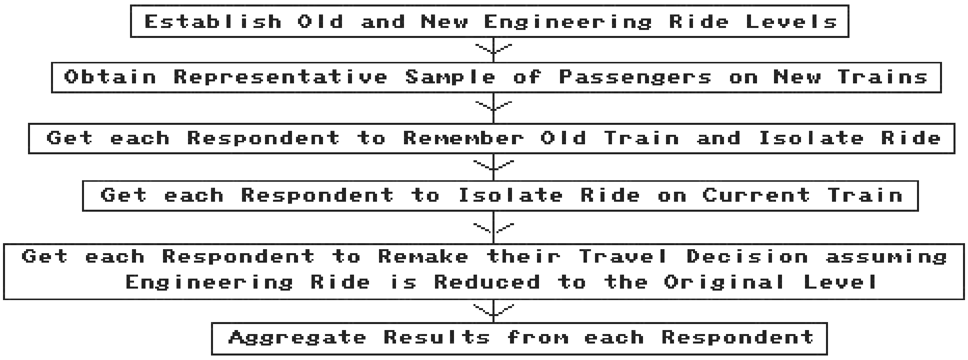
This approach estimates the value of a reduction in ride quality, by interviewing passengers on new trains which give an improved ride. Passengers are asked to remake their decision to travel today, but with the previous level of ride. The proportion of current passengers that would no longer make today's trip, would allow an estimate to be made of the value of ride.

Each respondent is asked to consider the effect of ride on one specific trip (rather than on their trip rates) as this makes the exercise simpler and less prone to error. It could be argued that considering only one trip per respondent may give an unrepresentative result. But as a large number of respondents' trips are to be considered, over different periods, any individual biases should cancel out. To ensure no overall bias exists, peculiar situations (such as late trains) are avoided.

5.2. Details of the Technique:

A number of tasks have to be carried out with this approach, they are shown in the flow diagram below. The details of each task and the difficulties associated with it are discussed in the order in which they are presented in the diagram.

FIGURE 5.5: VALUATION AFTER CHANGE IN STOCK.



5.2.1. Relationship Of Results To Engineering Ride:

A great advantage of this (and the following) approach/es is that the engineering ride levels have been directly experienced by all respondents. There is thus no need for elaborate descriptions to indicate levels of engineering ride. This means that any results will be directly related to measures of engineering ride.

Engineering ride would be measured over a series of runs for both old and new stock, for every section of a route. The levels of engineering ride associated with any respondent's journey could thus be estimated. A line therefore has to be selected before the complete transformation, so that measurements can be taken of the old stock (unless these are already available). Many lines have had a mixture of trains in the past which could make it difficult to establish the, "Old" level of engineering ride. To minimise this problem a Provincial route has to be selected, where new Sprinters have recently replaced a homogenous fleet of old D.M.U's or loco-hauled stock.

As this technique is based on a single rolling stock replacement, only two levels of engineering ride can be considered during a set of interviews. It is unlikely that more than one set of interviews can be managed in the time available, so the linearity of ride values cannot be investigated with this approach.

5.2.2. Obtaining A Sample Of Respondents:

The sample is taken from passengers travelling on the new trains. Passengers are first asked how much experience they have of the previous service. If their experience is inadequate, passengers are excluded from the survey. As such passengers could not give an accurate valuation of the change in ride.

The Centre for Transport Studies (1989) found that: 76% of passengers interviewed on a replacement Sprinter service had used the previous service, 6% of the sample having used it only once in the last year. So most people will have used the previous service (even if very infrequently) and thus should be able to produce some form of judgement.

Inaccuracy is likely to be more pronounced, if respondents are regular users of another rail service, which may interfere with their memories of past journeys on the current one (Baddeley (1982)). It is thus clear that previously infrequent users, whom this approach is designed to represent, may have difficulty with the task. This could mean their results are of limited value, reducing one of the main benefits of this approach.

The exclusion of such inexperienced passengers will not affect the results, unless this group have different values of ride from the rest of the passengers. This possibility is now considered.

A few respondents will have never used the service before. Most of these respondents are likely to be people who have changed their travel patterns: for example, people who have moved into the area recently, changed jobs or are making a one-off holiday trip etc. There is no reason to believe these people will have different values of ride from those people included in the exercise.

Some people will not have used the previous service, even though their travel patterns were compatible with it. Part of this group may have lost access to other modes, forcing them to use the new rail service. This part of the group will again be assumed to have ride values representative of the rest of the passengers and can therefore be ignored.

For the remainder of the group to be using the current service, they must have considerably improved their perception of the service, which suggests a peculiar value of ride. But the remainder of the group will not have based their travel decision on experience of the old stock, they are therefore likely to produce unreliable information. They will either have to be ignored, or be asked about their perception of the old service. The remainder of the group is likely to be small and their perception of the old service affected by contact with previous users and experience of other services. So their decisions may be best investigated by looking at the results of other passengers. This remainder of the group will therefore have to be ignored.

5.2.3. Recalling And Isolating Ride:

Respondents would be required to recall their experience/s of previous trains and isolate ride quality. Respondents would also have to isolate ride quality on the current stock to enable the two to be compared. Both these procedures would suffer from the interference effects described earlier (section 3.3).

By studying the effects of a coach refurbishment programme, M.V.A. (May 1986) have produced an indication of the size of newness effects (chapter two). Their results could be used to compensate for any Newness effect in a ride valuation approach. But some attributes were changed slightly between the trains they studied, meaning that Halo and Contamination effects would have also been incorporated into their results. The magnitude of the Newness effect would depend on how recently the stock was introduced (and probably to some extent, the age of the previous stock). M.V.A. do not say how recent the refurbishment was and give no idea of the rate of decay of Newness effects.

Although valuing ride on a route with cascaded stock could reduce Newness errors, this practice is no longer common on Provincial routes. Some attempt therefore has to be made to estimate the size of interference effects, so that any ride values produced can be adjusted accordingly.

Interference effects may be reduced by comparing: the new train against the new train with the old ride - rather than simply comparing the ride on the old and new trains. This would be achieved with a question like, "If this train rode like the previous one...". Even so interference effects would still probably accentuate the value of ride, as passengers may associate the old ride with other un-desirable features of the old train and subconsciously include these features into their calculations.

Care must be taken with the choice of older stock, as it is likely to be neglected causing an over-valuation of the improvements with new stock. It is possible that interference effects will be lower with Provincial trains because of some poor features of modern stock like: fewer toilets, less luggage space and poor/cramped seating layouts.

This approach suffers as respondents may have difficulty in accurately recalling the old stock (discussed earlier in section 4.7). To minimise this problem, we should do the experiment as close as possible to the full introduction of the new stock. But the early days of a new service are often characterised by a lack of awareness and teething troubles. So the later the exercise is done, the more likely it is that the full patronage increase has occurred and that an adequate number of previous non-users are represented on the current service. It would be pointless asking passengers whether they would still make today's trip, with ride reduced to the previous level, if these passengers had expected to make today's trip on an old train.

All passengers should therefore have made their original decision to travel today, on the basis that they would be using the new trains. Such a situation could take six months or a year to emerge. Even then a few passengers will have expected to be travelling on an old train and should therefore show no response to a reduction in ride to the original level. For the results to be applicable to other services the proportion of such people in the sample should be representative of that found on a stable service. Finally, the longer we wait after the introduction of the new stock, the smaller the Newness effect. If this approach is to be used, some point has to be found to satisfy these various trade-offs.

5.2.4. Replicating The Original Travel Decision:

Respondents would be asked to think back to when their decision to travel was made, hypothetically giving them time to plan for alternatives. As commuters and habitual travellers are unlikely to make a decision for every trip, they will be required to think back to when they decided to buy their season ticket etc. Respondents would then be asked to remake this decision, but with the current train having the same ride quality as the old train. The proportion of respondents who would no longer make today's trip, should give us an indication of the value of ride.

When this approach was first developed (chapter three) it was thought that respondents should not be directly asked about the change in ride quality between trains as this infers that ride quality has changed. Such inferences could give rise to errors (Guilford (1954)) where respondents feel that they have to give a value for ride even though it may be unimportant to them. Such an error would produce an excess value of ride.

However it was found in the depth interviews (chapter four) that most people who did not volunteer a change in ride, still considered there to be some difference. Eliminating people who do not volunteer a change is thus likely to produce an underestimate of the value of a ride change. It is therefore proposed that all respondents be asked to value any difference in ride between trains: this will be quicker and simpler than using a filter question without increasing error.

It was mentioned earlier, that the environment in which a decision is made may have an important effect (section 3.2). As the questions will have to be asked on-train, this means that we are not in the correct decision environment. As the individual is already making the planned journey his perception of rail travel may have changed from when he made the decision to travel. The original decision process may not therefore be exactly replicated.

With this approach each ride assessment is made under different conditions. The assessment of the old train will be based on memories of previous experiences, which to some extent will have been contaminated by other similar experiences. The assessment of the new train may be hyper-sensitised and heavily influenced by the current trip. It is not entirely clear how this will affect the results.

5.2.5. Applying Results To The Population:

Once the original travel decisions have been remade with the different level of ride, we want to be able to apply these individual results to the population. A range of valuation estimates for the population can be found by generating standard error statistics.

It is important to consider how well the valuation sample represents the population, as a poor representation could produce biased results (see section 3.1).

6. ACTUAL CHANGE IN RIDE - STOCK CHANGE - STATED PREFERENCE - CROSS-SECTIONAL:

6.1. Basic Issues:

This approach needs a number of stock types, with different engineering ride levels, running together in the same area. The most likely cause of this is where new stock has been only partially introduced on a service. It is unlikely that more than two significantly different stocks could be found in any Provincial area, meaning that only two levels of engineering ride can be considered with this approach.

Only people who have experience of both trains would be interviewed. It has to be assumed that those who are excluded from the sample have similar values of ride to those who are included. This is a reasonable assumption as most passengers will not know whether they will be travelling on a new or old train. As with the stock-replacement approach all respondents will be asked to value the change in ride.

Engineering ride would be measured over all sections of the route (in the same way as the after-replacement approach, section 5) allowing a close relationship between any ride values extracted and engineering differences.

Respondents would produce a valuation for the ride change between trains by playing some form of trade-off game. As with the after-replacement approach, respondents would choose between the new train and the new train with the old ride to minimise interference effects. The ride change would be traded against some known attribute, probably a change in fares.

Whether this approach is applied on a line undergoing stock replacement, or where stock is permanently running together, respondents' experiences should be more recent than with the old stock in an after-replacement approach. But with different stock running together, respondents are likely to be using more than one type of train in their area. Respondents will thus have made less use of each train, reducing the reliability of their assessments and making it more difficult to find a sample with adequate experience. Also having recent experience of a number of trains could increase the likelihood of recall interference (Baddeley (1982)).

6.2. Choice of Rolling-Stock Contrast:

There are a number of issues to be considered, in choosing between a replacement scenario and a situation where two types of trains (of similar age) are permanently running together.

Assessments of differences between two trains of similar age, will not suffer from Newness effects. With a replacement scenario, Newness effects may be greater than with the after-replacement approach (section 5) as differences will be highlighted with both stocks running together. As mentioned earlier, older stock is likely to be neglected causing a further over-valuation of the improvements with new stock.

Respondents may have more difficulty distinguishing between two trains of similar age, than between old and new trains. It may also be difficult to find two (or more) train types permanently running together that have significantly different levels of engineering ride. Where two or more types do permanently run together, it is likely that each stock is rostered according to some pattern (for example, the more comfortable trains may run the fast services). It is thus possible that respondents are experienced in only one of the types running on the line. Comparing stock designed for different journey types is also likely to increase Halo and Contamination effects.

For these reasons it has been decided to continue with the replacement scenario alone.

People could be interviewed on old or new trains. Obviously it would be unwise to interview on both as the results would not be comparable. It is perhaps best to interview on new trains as most passengers on the new trains will have experience of the old trains. A much smaller proportion of passengers on the old trains are likely to have experience of the new trains, making sampling difficult.

6.3. Choice of Survey Location:

At this point we have to decide whether to develop an on-train or household technique. A summary of the factors behind the choice (discussed in detail in chapter three) is given below.

A household method provides: the correct decision environment, a sample of infrequent users and produces no hyper-sensitivity effects.

An on-train method provides: a captive audience, makes questions appear more relevant and thus interesting, is an appropriate environment - aiding recall and easily produces a sample of respondents with sufficient experience to accurately answer questions - thus saving time.

Whichever method is selected the sample will be biased towards experienced passengers, as they will be the only ones capable of giving an assessment of the ride on the new stock. Few infrequent users will have tested the new stock and so will be under-represented, even in a household survey. As mentioned before, we need to find whether infrequent users are differently critical of ride. It is also important to establish the size of the effects of hyper-sensitivity. The effect of the decision environment is crucial and thus has to be investigated before a decision can be made.

The above three factors represent the only advantages of a household approach and if no differences are apparent, it would be most beneficial to conduct the survey on-train.

7. ACTUAL CHANGE IN RIDE - SUSPENSION/TRACK - STATED PREFERENCE - CROSS-SECTIONAL:

7.1. Introduction:

With this approach, passengers ride and assess one current stock, over different qualities of track. To achieve a valuation of ride in this way we must find a sample of the population who have experience of the required track sections.

There are two ways of satisfying this requirement. We can get people to compare the ride on different trips they have made using the same stock. Alternatively we can ask people who are currently travelling, to compare the ride on sections of this trip.

7.2. Choice of Technique:

Choosing between the two techniques outlined involves similar arguments to those considered earlier, when the most effective ways of establishing a textual ride scale were considered (section 4.7.6). The relevant points are outlined below - for detail see the earlier section.

Considering sections of the current trip will produce order effects, reducing the accuracy of the individual assessments. It is likely to be easier for respondents to compare ride between different trips than sections of one trip (for example, Guilford (1954)). But it would be difficult to find respondents with experience of a number of trips using the same rolling stock, that have significantly different levels of engineering ride. A large number of people must have experienced the same trains on the same routes, to enable their responses to be aggregated. Such a situation is unlikely to be found. With the current-trip approach, we are at least sure that everyone has recent experience of the required sections.

A multi-trip approach assesses more distant memories than the single-trip technique. As ride quality is not a major issue, respondents are unlikely to retain much detail and information will be compressed and difficult to categorise (Baddeley (1982)). There is therefore likely to be significant interference with such specific recall. As the relevant ride experiences will be similar (being related to the same stock) there will be few cues to aid in the recall of ride on a particular section.

Asking about each section of a trip (with the current-trip technique) could lead to hyper-sensitivity, unless all assessments are done at the end of the relevant track sections resulting in increased recall error.

With a multi-trip technique ride may not be the only difference between services. Even if both lines have the same stock, it may be that one journey is associated with, say, less reliability - this could bias any value extracted.

With a multi-trip technique we must make a large number of engineering ride measurements to account for all the experiences of passengers. We cannot be sure that all respondents' experiences relate to the current track standard.

Although these arguments seem fairly evenly balanced, the multi-journey technique's need for very experienced respondents makes this approach less likely to succeed. The current-trip technique is therefore the only track-based valuation approach to be developed further.

7.3. Current-Trip Technique:

With the current-trip technique passengers would be interviewed along a route representing a variety of track standards, so that different levels of ride are experienced in the same stock. With this technique engineering ride would be measured throughout the route. The route would then be divided into sections that would probably correspond to whether trains were running on or off main lines.

Average engineering ride would be derived for each section. As an example, with the Weymouth to Westbury service, the route uses a Network South East mainline from Weymouth to Dorchester West, Provincial track from Dorchester West to Castle Cary and InterCity track from Castle Cary to Westbury. At the end of the route passengers would be asked to trade-off the various levels of ride they had just experienced against a known attribute, like fares.

As ride is being compared between two identical trains there would be minimal interference effects. But for this technique to be successful, people will need to accurately detect the small differences in engineering ride between each track section. The difference between track sections is likely to be smaller than is immediately apparent as the worst track will be traversed at lower speeds.

The difficulty of finding a route and the need to make the respondents' task as simple as possible, means that three is probably the greatest number of sections that can be considered. Interviews would take place at the end of the last section to avoid hyper-sensitising respondents.

The most significant problem with this approach is the existence of order effects and recall interference. The succession of jolts etc. experienced by each respondent are bound to be amalgamated into an overall picture of the journey (Baddeley (1982)). As these experiences have occurred under very similar conditions this interference is likely to be significant. It may be very difficult for respondents to recall and isolate the events of one part of a journey and compare these with the events of another part of the same journey. Furthermore it is possible that respondents' mental subdivisions of the route, are different from those we are asking them about. If this were the case it would be even more difficult for respondents to isolate experiences.

As stated earlier it is important to ensure that each respondent has ridden over the whole of each section specified: so that they everyone is valuing the correct engineering ride levels. To minimise interference it would also be beneficial, if each respondent was only travelling over the relevant sections. However this would make sampling difficult and time consuming, even on a route with substantial through traffic.

8. CONCLUSIONS AND IMPLICATIONS FOR LATER RESEARCH:

8.1. Introduction:

In this chapter the ideas generated in earlier parts of the research have been developed. In particular we have produced a textual ride scale based on information from previous psychological research and the depth interviews described in chapter four. The proposed scale is a continuous graphic scale with a vertical orientation and an anchor at either end. The ride valuation approaches (originated in chapter three) have also evolved in the light of information produced by the depth interviews.

It is clear from the discussion in this chapter that further information is required before the final selection and implementation of valuation approaches can take place. These needs and the ways of fulfilling them are now discussed.

8.2. Developing Textual Scale Locations:

The main need at this stage is to develop a set of textual ride scale locations that can be valued later. This is best achieved by getting respondents to assess a number of specified sections of their current trip. These repeated assessments allow a number of issues to be investigated (section 8.3). Assessing the sections of one trip is considered the most effective dependent design (see section 4.7.6). With this design, individuals' results are comparable and amenable to aggregation. As stated previously (section 4.7.6) the use of this design does not preclude the development of independent textual locations later.

Engineering ride is measured throughout the route to enable its relationship to the textual locations to be established.

8.3. Other Information:

The other approaches, for establishing a value of ride, are also considered in the design of this intermediate stage. The issues that need to be included are now considered.

8.3.1. The Effects Of Personal And Other Characteristics:

Some types of people and trip could be more critical of ride than others. Each respondent is therefore asked for information about themselves and their trip. This should include: age, sex, use of car, origin-destination (trip length), trip purpose, regularity of rail use, whether they are standing and how often they have to stand when using rail generally.

8.3.2. The Effects Of Hyper-Sensitivity:

An effective way to test the effects of hyper-sensitivity is to inform some of the sample that they will be assessing ride during the current journey, but not the rest.

One group of respondents is therefore given questionnaires at the beginning of the sections to be assessed, rather than at the end. This group thus rates ride while passing over the relevant sections - causing them to be hyper-sensitised.

8.3.3. The Effect Of The Decision Environment:

Initially it is important to ask where respondents were when they made the decision to travel. If the environment does appear to significantly affect decisions, the valuation can then be conducted in the appropriate place.

To establish the importance of the decision environment, respondents are first asked to rate the importance of ride on-train. The same respondents then produce a similar rating at home using a mail-back questionnaire (risking a low response rate). By the time respondents produce their second rating they should have forgotten the initial one.

8.3.4. The Consistency Of Respondents' Ratings:

To investigate the consistency of individual's estimates, relative to the population's estimates. Respondents must be asked to assess more than one level of engineering ride, this is a major reason for the choice of a dependent design (section 8.2).

8.3.5. The Distribution Of Perceptions For A Given Engineering Level:

No special effort is needed to gather this information. Positions on the textual scale produced by each respondent, for given levels of engineering ride, are all that is required.

8.3.6. The Ability To Detect Small Differences In Engineering Ride:

It is important to establish whether individuals are able to distinguish between the engineering ride levels of various track sections. Ideally respondents should be able to do this at the end of their trip. This is investigated by asking respondents to assess specific sections of the route they have just passed over. The inclusion of a don't know response further indicates respondents' abilities.

It is thought that individuals' overall ride assessments are more reliable than those of certain sections of a trip. Respondents are therefore expected to produce assessments of the overall ride on the train so far.

8.3.7. Level Of Respondents' Experience:

An indication of peoples' ability to compare different trains and routes can be found, by asking respondents about the local rail services used in the last year.

8.3.8. Linearity Of Ride Valuation:

The linearity of ride valuations is important as it may affect the level of error with the hypothetical technique. Linearity will also the affect the application of any ride values produced.

Linearity cannot be investigated until ride values have been extracted. By then it is too late to affect the decision whether to use the textual scale approach (section 4.5.2). Development of the textual scale approach will nevertheless continue and if valuations do appear significantly non-linear its results will be treated more cautiously.

8.3.9. Ride Values Of Previous Infrequent/Non-Users Different:

We need to establish whether the ride values of infrequent and non-users are significantly different from those of other users. If there is a difference, infrequent and non-users must be taken account of in the valuation survey design.

Again this cannot be investigated fully until the valuation exercises have been done, which is too late to affect the choice of valuation approach/es. However the importance of ride will be investigated, to establish the effect of the decision environment (section 8.3.3). This when related to the frequency of rail use should give a good indication of any differences with infrequent and non-users.

8.4. Interview and Sample Design:

The combination of the tasks considered above means that the main interviews have to be conducted on train. It is important that a line is chosen with several discrete sections characterised by different levels of engineering ride. Most passengers should be travelling over all these sections, ensuring that they have sufficient experience for the exercise. The implementation and results of these interviews are discussed in the following two chapters.

Chapter Six Ride and Other Attributes

1. INTRODUCTION:

1.1. The Opportunity for Further Analysis:

At this stage of the work an opportunity arose to take part in a large scale British Rail (Provincial) passenger survey. This exercise was to investigate the change in passengers' perceptions, as a result of the introduction of new rolling stock. The analysis of this data for the Provincial contract generated some information on perceptions of ride, but detailed analysis was beyond the scope of the Provincial brief. Further analysis is therefore done (as part of this research) as this may provide some answers to the issues raised in chapter five.

This study was not specifically designed with the Ph.D. in mind and as a result the questionnaire used (appendix four) does not completely follow the guidelines laid down in the earlier chapters. Nevertheless the opportunity was taken to incorporate this large data set in the thesis as it would provide an opportunity to validate any results produced by the structured Ph.D. interviews (described in the next chapter) and may also provide further insights into the ride issue.

1.2. Survey Detail and Structure of Analysis:

Two sets of surveys were conducted on the Bristol-Weymouth service (see appendix three). The first survey was done on modernisation plan D.M.U.s, while the second was done on replacement Sprinters. Neither of these groups of rolling stock were completely homogeneous, though the old D.M.U.s were virtually indistinguishable from each other. There was a greater difference between the two Sprinters considered (the class 155 had improved layout and decor compared to the class 150/2). The old D.M.U.s were therefore treated as one group during the analysis, while the two types of Sprinter were treated separately.

The British Rail data sets are very large (900-1,000 subjects) and were collected over seven/eight days. The sample is considered a reasonable representation of the passengers on the line. This survey (unlike those specifically designed for the Ph.D) covered both short and long distance travellers.

The analysis conducted in this chapter falls broadly into two main sections. Initially general findings from the data, for example assessments of ride quality, are considered. In the second stage of the analysis the relationships between ride assessments, personal/trip characteristics and the ratings of other service attributes are investigated.

2. GENERAL RESULTS:

Various attributes were rated by respondents on a five point scale. Ride achieved the results shown in the table and graph below.

TABLE 6.1: ASSESSMENTS OF RIDE QUALITY.

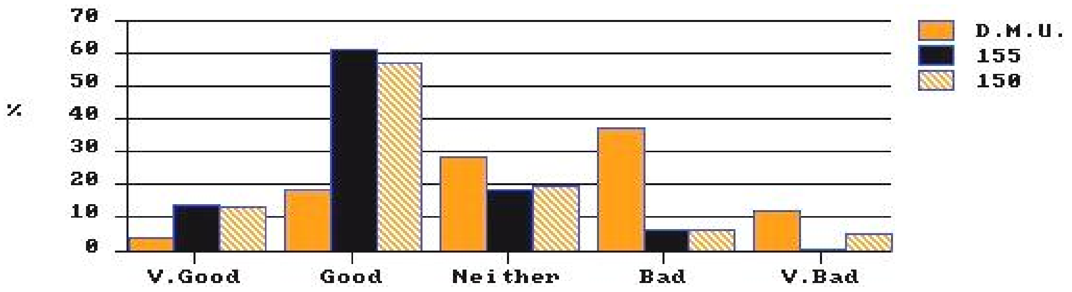
OLD D.M.U.	FREQUENCY	VALID %	CUMULATIVE %
Very Good	31	3.6	3.6
Good	162	18.6	22.1

Neither	249	28.6	50.7
Bad	324	37.2	87.8
Very Bad	106	12.2	100.0

155 SPRINTER	FREQUENCY	VALID %	CUMULATIVE %
Very Good	119	13.6	13.6
Good	535	61.2	74.8
Neither	159	18.2	93.0
Bad	56	6.4	99.4
Very Bad	5	0.6	100.0

150 SPRINTER	FREQUENCY	VALID %	CUMULATIVE %
Very Good	11	13.3	13.3
Good	47	56.6	69.9
Neither	16	19.3	89.2
Bad	5	6.0	95.5
Very Bad	4	4.8	100.0

FIGURE 6.1: PASSENGER ASSESSMENTS OF RIDE QUALITY
BRISTOL TEMPLE MEADS - WEYMOUTH (1989).



It is clear from the above that the ride on the old D.M.U.s was not considered very highly; "Bad", was the most commonly selected category. The mode for both the Sprinters was, "Good", little difference was apparent between the two Sprinters (the 155 should be slightly rougher than the 150/2). In order to test for differences in the perceptions of ride quality between the trains a Kruskall-Wallis non-parametric (because of the ordinal nature of the data) ANOVA. was used. A null hypothesis was constructed that, "There are no significant differences between the perceptions of ride on the various trains".

Kruskall-Wallis Statistic (corrected for ties) = 541.853

Significant @ 99.9% with 2 d.f.

The result clearly shows a difference in perceptions of ride quality between the various trains. The pairs of trains exhibiting differences can now be found using an adaptation of Scheffes technique of paired comparisons.

TABLE 6.2: COMPARISONS OF RIDE PERCEPTIONS BETWEEN TRAINS.

	155	150/2	Old D.M.U.
155	-	-	-
150/2	N/S	-	-
Old D.M.U.	99%	99%	-

From the table above it is clear that no significant difference emerges between the Sprinters, even with this large sample size (903 Class 155, 83 Class 150/2). Both the Sprinters are perceived to be significantly smoother than the old D.M.U.

The ratings of the attributes before and after the change can be seen in the tables below. The line across the tables represents the point where ratings changed from negative to positive. Ride was originally considered the third worse attribute in the study, better than only noise and catering. Ride's position improved to eighth with the introduction of the Sprinter service. In fact smoothness of ride is perceived to be the largest change to result from the introduction of the new trains. The introduction of the Sprinter service resulted in a significant improvement in the scores of all attributes, apart from view and luggage space (Centre for Transport Studies (1989)).

TABLE 6.3: ASSESSMENTS OF ATTRIBUTES - OLD D.M.U. SERVICE.

ATTRIBUTE	RANK (Worst first)	NOTES
Catering	1	None Provided
Noise	2	
Ride	3	
Toilet	4	Presence not always known
Cleanliness	5	
Seating	6	
Luggage Space	7	
Ease of Operating Doors	8	
Ability to Read	9	
Heating/Ventilation	10	
Lighting	11	
Ease of Entry	12	
View	13	

TABLE 6.4: ASSESSMENTS OF ATTRIBUTES - SPRINTER SERVICE.

ATTRIBUTE	RANK (Worst first)	NOTES
Catering	1	None Provided
Luggage Space	2	
Noise	3	
Toilet	4	Presence not always known
Cleanliness	5	
Seating	6	
Lighting	7	
Ride	8	
Ability to Read	8	
Ease of Entry	10	

View	11
Ease of Operating Doors	12
Heating/Ventilation	13

3. RELATIONSHIPS WITH RIDE ASSESSMENTS:

3.1. General Issues:

Previous research (described in chapter two) suggests that individuals' assessments of ride quality may be affected by the levels of other related attributes, like noise. Such relationships could have an important bearing on any findings produced by this investigation. In this section we therefore consider two groups of such relationships: firstly between ride assessments and personal/trip characteristics, secondly between ride assessments and the ratings of all other stock attributes.

For each relationship a number of cross-tabulations are produced. The cross-tabulations are tested statistically using Chi-Squared and (where applicable) Kendall's Tau-C rank correlation coefficient. Non-parametric measures are selected because of the nominal and ordinal nature of the data. Even with samples of around 900, some of the initial Chi-Squared results are invalid, as too many cells have expected values below five and/or the minimum expected frequency is less than one (Groebner and Shannon (1985) state that 80% of cells should have expected frequencies of greater than five and that all cells should have expected frequencies greater than one).

All relationships are examined twice: for the old D.M.U. data and also the 155 Sprinter data. As there are significant differences between the Sprinters, the class 150/2 respondents (the smallest group) are excluded from the analysis. All references to Sprinters that follow, therefore refer exclusively to class 155 units. This means that any variation in ratings is almost entirely a result of differences in perception between individuals.

Large sample sizes mean that even weak relationships are statistically significant. However, closer examination reveals that although most variables do exhibit some relationship with ride, few of them can be considered strong. The large number of significant results thus give an exaggerated impression of the relationships between ride assessments and those of other attributes. Each relationship is considered in detail below, 3-D barcharts were used to help pick-out any particular trends.

The following analysis is divided into three sections. Initially we consider the connections between ride assessments and personal characteristics; then we consider both the univariate and multivariate relationships between ride assessments and those of other attributes.

3.2. Relationships with Personal Characteristics:

3.2.1. General:

As a result of the statistical problems mentioned in the previous section, some response categories have had to be merged. Care has been taken to combine the most similar categories, to minimise the loss of meaning that occurs with such techniques. Even after merging the Sprinter ride assessments are so bunched, that most of the Chi-Squared results are invalid.

3.2.2. Ride By Purpose:

OLD D.M.U. CHI-SQUARE 98.93% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID.

To establish valid results it was necessary to merge the employers business category with forces duty. Holidays and visiting friends/relatives were also merged. Even after merging the Sprinter results were still invalid and so no conclusions can be drawn from this data.

Work, employers business (including forces duties) and education trips displayed similar results, being the most critical of ride. Shopping trips represented the other extreme (being considerably less critical). Personal business, day out and holiday trips (including visiting friends/relatives) were also less critical than the work etc. categories - falling between the two groups.

From this result there appears to be a distinction between discretionary and non-discretionary trips. It seems that passengers making work/education trips are more critical, than those on trips of other purposes. It may be that passengers who have to make such journeys regularly, become less tolerant of ride.

3.2.3. Ride By Sex:

OLD D.M.U. CHI-SQUARE 96%, TAU-C 99.14% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.76% SIGNIFICANT.

With the old D.M.U. data males are shown to be slightly more critical of ride quality, this was backed up by the Sprinter data. But sex produces only weak correlations (Old D.M.U: -0.089 and Sprinter: -0.097) and in the 3-D plots the relationship was not very clear.

3.2.4. Ride By Age:

OLD D.M.U. CHI-SQUARE 99.99%, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

With the old D.M.U. data the oldest two categories are much less critical of ride quality than the (two) 16-44 categories. The 45-59 and under 16 groups fall between these extremes. The Sprinter data also suggests older passengers are less critical, though the result is not as clear. The negative correlations produced are (Old D.M.U: -0.172 and Sprinter: -0.124).

The lower sensitivity of older people does seem to follow the experience of research into travel sickness, which is less common with ageing (for example, International Standards Organisation (1985)). But this should make the under 16 group very critical of ride. The under 16 result may be related to experience of other modes of travel, this group would have less experience of alternative modes of transport, which may lower their expectations. If the ride became much rougher, the elderly would be expected to be relatively more critical, because of the need to brace themselves against jolts etc.

3.2.5. Ride By Job:

OLD D.M.U. CHI-SQUARE 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID.

In an attempt to establish valid results, it was necessary to merge the Keeping House and Unemployed categories. Ideally more obviously similar categories would have been merged, but this was the most effective way to remove the two small categories that were causing statistical difficulties. Even with this modification the Sprinter results were still invalid and so no conclusions can be drawn from this data.

Retired people prove less critical than the others, this may be related to older respondents being less critical (section 3.2.4). The Unemployed (including Keeping House) are also less critical - though not to the same extent. Full Time workers are the most critical of ride. These results may be related to those of trip purpose (section 3.2.2).

3.2.6. Ride By Travel Regularity:

OLD D.M.U. CHI-SQUARE 99.85%, TAU-C 99.84% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 88% SIGNIFICANT.

There is an indication here, with both the old D.M.U. and Sprinter results, that more frequent users are more critical of ride; a very slight positive correlation (Old D.M.U: 0.081 and Sprinter: 0.029) is observed. This trend may tie in with the earlier observation that non-discretionary trips may be more critical (section 3.2.2).

3.2.7. Overview:

Initially it appears that passengers on work, business and education trips are more critical of ride than passengers on other trips. Generally people in employment were more severe in when rating ride. There is an indication that males may be slightly more critical than females. Older people tend to be less critical - this lower sensitivity corresponds with travel sickness research (British Standards (1987), International Standards Organisation (1985)). Finally there is a suggestion that those who use the train most frequently are more critical.

Closer examination of the relationships just discussed suggests that there may be some underlying factor. Most of the personal/trip characteristics are intuitively related to each other as well as with assessments of ride quality. For example: middle aged people are most likely to be employed and are thus likely to be making work trips. Those making work trips will travel more frequently and are more likely to be male.

It is therefore clear that none of the characteristics considered in this section can be thought of as truly independent. This lack of independence suggests that some personal/trip characteristics may not be directly related to assessments of ride quality. One of these characteristics alone could be responsible for all the variations in ride assessments recorded above. Correlation does not necessarily indicate causation - each relationship recorded in this section could be a proxy for other indirect relationships. For example: a recorded relationship between ride assessments and frequency of travel could in reality be generated by differences in trip types.

Such inter-relationships make it very difficult to disentangle any underlying effect. Any attempt to do this statistically is made difficult because of the nominal nature of two of the variables (purpose and job). What is clear however, is that all these relationships point towards assessments of ride being more critical on suburban lines - characterised by frequent work trips and mainly middle aged, male passengers. It should be noted that commuters may not value ride any more highly - they may be less tolerant about everything. Such relationships would need to be taken account of, if any ride valuations are to be widely applicable.

A further issue to arise from this discussion is perhaps the redundancy of the job question in the ride valuation exercise. It is always difficult to ask members of the public about their employment and other variables, particularly trip type, do give a reasonable indication of whether someone is employed or not. These other characteristics are thus likely to act as a proxy for employment.

3.3. Univariate Relationships with Other Stock Attributes:

3.3.1. General:

In this section relationships between ride and other attributes are examined, both where connections are expected from previous research and where they are not. Both forms of relationship are tested to enable comparisons with previous research and to investigate the possibility of interference effects - even from apparently unrelated attributes.

A main issue throughout this research is the possibility of ride quality ratings being interfered with by other attributes, for example noise. We want to establish the extent to which differences in non ride attributes affect peoples' assessments of ride quality. From this we can estimate the amount of interference in any ride values eventually produced. We may be able to get an indication of such relationships by comparing individuals' perceptions of ride with their perceptions of other service attributes. Thus we can tell whether people who consider a particular attribute bad, tend to consider ride to be bad and vice-versa.

It must be stressed that the results of such an analysis may not be the same, as a situation where ride assessments are taken while the engineering levels of other service attributes are varied. The correlations based on these perception relationships may not therefore indicate interference. Nevertheless an analysis of perceptions enables an investigation of far more relationships than would be possible in any real situation.

An attempt is therefore made to gauge interference effects using such an approach in this (univariate) and the following (multivariate) sections. In this section the six expected relationships are examined first.

3.3.2. Ride By Assessment Of Seating:

OLD D.M.U. CHI-SQUARE 99.99%, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

There is a strong relationship between individuals' ride assessments and their assessments of seating quality. Positive correlations of (Old D.M.U: 0.321 and Sprinter: 0.238) were observed. It was also observed that the higher passengers rated seating (Old D.M.U) the more spread were the ride assessments. With the Sprinter those who rated seating, "Very bad" exhibited more spread in their ride assessments.

This relationship would be expected from the relevant British and International Standards (British Standards (1987), International Standards Organisation (1985)). It may be that one way of slightly improving perceptions of ride quality would be to improve seating.

3.3.3. Ride By Assessment Of Heating/Ventilation:

OLD D.M.U. CHI-SQUARE 99.99%, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

A further strong relationship between individuals' ride assessments and their assessment of heating quality. Positive correlations of (Old D.M.U: 0.297 and Sprinter: 0.262) were observed. Again it was found that those who rated heating positively (Old D.M.U) or rated it, "Very bad" (Sprinter) showed more spread in their ride assessments. This relationship would be expected from the British and International Standards (British Standards (1987), International Standards Organisation (1985)). Again it may be possible to slightly improve perceived ride quality by improving heating.

3.3.4. Ride By Assessment Of Noise:

OLD D.M.U. CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

This is the strongest relationship observed, with positive correlations of (Old D.M.U: 0.528 and Sprinter: 0.349). With the Sprinter data, those who assessed noise the most critically exhibited more variation in their ride assessments. This relationship would be expected from the British and International Standards (British Standards (1987), International Standards Organisation (1985)). Looking at the 3-D graph, much less spread is visible than with the other relationships.

The closeness of this relationship means that care must be taken when trying to isolate ride quality from noise. This relationship means that an effective way to improve perceptions of ride quality, may be to improve sound insulation on trains. It may also be cheaper to reach a certain perceived ride level, by improving sound insulation rather than improving the suspension of the train. Although respondents in the depth interviews seemed able to isolate noise from ride, it seems that subconsciously they may be less able to make such a clear division. More care will therefore have to be taken in holding noise constant between ride assessments.

3.3.5. Ride By Assessment Of Lighting:

OLD D.M.U. CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

A clear positive relationship, with correlations of (Old D.M.U: 0.254 and Sprinter: 0.213). With the Old D.M.U. data, as respondent's light assessments got less critical their ride assessments became more variable. This relationship would be expected from the British and International Standards (British Standards (1987), International Standards Organisation (1985)).

3.3.6. Ride By Assessment Of Ability To Read:

OLD D.M.U. CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

A strong positive relationship, with correlations of (Old D.M.U: 0.305 and Sprinter: 0.209). With the Old D.M.U. results, those individuals who gave the ability to read as, "Good" or, "Very good" gave more dispersed ride ratings. The ability to read does not represent a single attribute, it is a combination of attributes. The ease of reading, while on a train, is obviously closely related to ride quality, this relationship would therefore be expected to show higher correlations than it did. There was some difficulty with the ability to read question (some interpreting this as a question about eyesight) which may account for this weakening.

3.3.7. Ride By Assessment Of View:

OLD D.M.U. CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.
SPRINTER CHI-SQUARE INVALID, TAU-C 99.99% SIGNIFICANT.

A positive relationship, with a correlations of (Old D.M.U: 0.125 and Sprinter: 0.146). Graphically the relationship is less clear than many of those previously discussed. With the old D.M.U. results individual's ride assessments are widely spread, except where view was considered, "Very bad" (which was not very often); with the Sprinter results those who considered view, "Very bad" exhibited most spread. Some relationship would be expected from the British and International Standards (British Standards (1987), International Standards Organisation (1985)).

3.3.8. Ride By Assessments Of Other Attributes:

Statistical tests were also performed to test the relationships between assessments of ride quality and those of other attributes - where a connection would not be expected. These tests were done to investigate the possibility of a Halo effect: where an overall impression of (un)favourableness colours a respondents view of the attribute under consideration. Tests were carried out on the following non-ride assessments: level of fares, frequency of service, reliability, cleanliness, luggage space, ease of enter/exit, ease of operating doors, toilet facilities and catering.

The large size of the data sets meant that every relationship proved very significant (nearly all reached 99.99%). Nevertheless all the relationships were very weak, displaying a maximum Tau-C correlation of 0.226. The results thus give some indication of a small Halo effect.

One feature of these results is that they suggest respondents were taking the exercise seriously. For example, the correlations between respondents' assessments of the level of fares and those of ride quality were negative. All other relationships exhibited positive relationships as expected. The far result is important as it tells us that respondents realised the good points of the ride and fare scales were at opposite ends. This implies that respondents were taking some care over their responses and did not just tick all the boxes down one side of the questionnaire.

In many cases respondents ride assessments were more spread where the other attribute was considered, "Good" or, "Very good". This suggests that people with strong opinions on a particular attribute were able, to some extent, to isolate their ride assessments from this.

3.4. Multivariate Analysis - General:

The data is further analysed using both factor analysis and multiple correlation. Multivariate techniques are used at this stage to enable an examination of the interaction of the assessments of non-ride attributes. It is clear from the previous section that many assessments are significantly (although weakly) related to ride assessments. It may be that some of these weak relationships are cumulative. In such a case a combination of non-ride assessments may exhibit quite a strong relationship with ride assessments. Such an enhanced relationship could suggest more significant interference than that implied by examining non-ride assessments individually.

As mentioned earlier this data is not wholly suited to parametric analysis, as we cannot assume that assessments are on an interval scale. The difference between, "Neither good/bad" and, "Good" is unlikely to be of the same magnitude as the difference between, "Good" and, "Very good". However so that some indication of the extent of interference with ride assessments can be gained, parametric correlation coefficients have been calculated. The following analysis should therefore be treated with caution.

3.4.1. Factor Analysis:

Factor analysis is carried out to investigate the way assessments of each attribute vary together. This multivariate technique thus allows us to identify the most similar assessments. The procedure is based upon a correlation matrix and attempts to group the most similar assessments into a factor. As assessments of ride will be placed into one of these groups, factor analysis further indicates those attributes that may be difficult to isolate from ride.

All factors are extracted using Principle Components Analysis. Factor solutions are then rotated using the Varimax method, to make the solutions more interpretable (for example, Norusis (1988)). With both the old D.M.U. and Sprinter data four factors are extracted by the analysis. The factor loadings (after rotation) can then be interpreted to find the attributes that vary together. The factors extracted from both data sets are broadly similar, cross-validating the results.

Each attribute is listed below (for both old D.M.U. and Sprinter data) under the factor with which it displayed the highest factor loading. The attributes are listed under each factor in the order of their factor loading (shown in brackets). The factor loading is a measurement of the closeness of an attributes relationship with that factor. The factor loadings can be interpreted as the standardised regression coefficients (that is between 0-1) or as an indication of the correlation between each attribute and that factor.

TABLE 6.5: OLD D.M.U. FACTOR ANALYSIS.

FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
Ride (0.8)	Doors (0.8)	Frequent (0.8)	Fares (-.7)
Noise (0.8)	Enter (0.8)	Reliable (0.8)	View (0.5)
Seats (0.6)	Luggage (0.6)		
Read (0.6)	Clean (0.5)		
Light (0.5)	Toilet (0.5)		
Heating (0.5)			

TABLE 6.6: SPRINTER FACTOR ANALYSIS.

FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
Ride (0.8)	Doors (0.8)	Fares (-.7)	Frequent (0.8)
Noise (0.8)	Enter (0.8)	Clean (0.5)	Reliable (0.7)
Heating (0.6)	Luggage (0.4)	Toilet (0.5)	
Light (0.5)	View (0.4)		
Seats (0.5)			
Read (0.5)			

From the factor analysis it is clear that ride assessments are consistently associated with the same five non-ride assessments: noise, heating, lighting, seating and the ability to read. Of these, noise assessments appear to be the most strongly associated with the ride assessments. In fact, ride and noise appear to be the most closely related pair of assessments (other than doors/enter and frequency/reliability). If a relationship between ride and noise assessments suggests a relationship between ride assessments and engineering noise levels, this finding reinforces the view that the isolation of ride from noise requires considerable care. Even so, it seems likely that any ride values produced in this research will be contaminated to some degree by noise.

The other four assessments associated with the ride assessments exhibit a more distant relationship. Though it may still be wise to consider these attributes during the valuation exercise.

Ride's loading in the other three factors is very low (below 0.2 for both data sets) suggesting that contamination from the attributes represented in these factors is much less of a problem.

3.4.2. Correlation Analysis:

It is clear from the foregoing analysis that passengers' ride assessments may not reflect ride alone - opinions of other attributes could have interfered with passengers' judgements of ride quality. There are two aspects to this interference: firstly there is the Halo effect where assessments are contaminated by an overall favourableness towards the object in question. Secondly, there is the inability to completely isolate attributes from each other. As the eventual aim of this research is to estimate a value of ride, it would be useful to consider how much of the ride ratings are truly a reflection of ride and how much is due to such contamination. If this can be achieved, the eventual ride values may be adjusted to take account of this interference.

In this section an attempt is made to estimate the size of the overall interference in the ride ratings by those of other attributes. From this it is hoped that some indication of what would happen to ride ratings, if actual levels of non-ride attributes were varied, can be given. It must be remembered at this stage (see section 3.3) that the relationships between assessments of ride and those of non-ride attributes, do not necessarily prove a similar relationship between ride assessments and engineering changes in other attributes. Nevertheless analysis of the relationship between ride and other assessments should give some indication of any link between attributes. An attempt to estimate an overall interference figure is therefore made.

The best way to determine the extent of interference is to look at the square of each assessments correlation with the ride assessment. The squared correlation coefficient indicates the proportion of the variation in the dependent variable (in this case the ride assessment) that is accounted for by any independent variable. Multiple correlations can also be produced and squared, indicating the proportion of variance in the dependent variable accounted for by a group of independent variables. Unfortunately this interpretation of squared correlation coefficients is not acceptable, when using the non-parametric measures of correlation as reported in the above cross-tabulations (for example, Hayes (1988)).

There are a number of factors that suggest any interference value produced from this analysis will be exaggerated. The parametric correlations produced are generally higher (about 29% above) than the non-parametric measures. As already stated the parametric measures will be less valid, in this context, than the non-parametric measures used earlier.

Further, the anchors used on the British Rail ride scale are not very specific, for example, "Very good". The language used in these descriptions could be applied to any attribute and this may encourage respondents to make more general assessments than is desirable. The descriptions of ride levels that will be given to respondents in the ride valuation exercises, will be developed solely for this purpose. We would thus expect assessments based on the vague British Rail rating scale, to show a greater degree of interference than any valuations produced during this research.

Conducting repeated attribute assessments, with the same scale is also likely to increase interference (for example, Guilford (1954)). In the valuation exercises ride would be the only (or one of few) attribute/s that is being assessed.

Finally, the presence of the ability to read ratings in the correlations will also cause interference to be over-estimated, as ride quality may be a significant component of this rating.

According to the correlation analysis, between 35% (Sprinter) and 55% (Old D.M.U) of the variation in ride ratings can be explained by the scores of other attributes. Noise explains 87% of the accounted variance (Old D.M.U) and 77% of the accounted variance (Sprinter). The most effective regression equations produced during the analysis for both sets of data had: noise, ability to read, seat and heating as the prominent independent variables. With both old D.M.U. and Sprinter models, adding unexpected variables made no difference to the explanatory power of the equation.

From this analysis, we can say that (from the average of the two data sets) the level of interference could be as high as 45%; though relationships between individuals' assessments do not necessarily equate to interference effects (section 3.3). If ability to read is removed from the equations, we would expect this figure to fall by 2-3%. We can therefore say that, in the worst case, ride assessments may be contaminated by approximately 42%. Of this 42% approximately 4-5% is a result of the overall favourableness of the train (Halo), while the remainder is a result of the difficulty of isolating related attributes.

4. SUMMARY:

Respondents clearly detect the ride difference between the old D.M.U.s and Sprinters (99% significant) though not between the 150/2 and 155 Sprinters. This gives some idea of the sensitivity of ride perceptions. The Sprinter's ride is generally seen as, "Good" while that of the old D.M.U.s is seen as, "Bad". Ride is the third worst attribute on the old D.M.U.s, improving to eighth on the Sprinters.

When looking at the relationships of ride ratings with personal characteristics and ride ratings with other attribute ratings, the very large samples mean that even weak relationships are statistically significant.

It is clear that those passengers making frequent, non-discretionary trips seem to be the most critical of ride quality. It was suggested earlier that trip purpose, age and job may be measuring the same thing to some extent; asking the difficult question about a respondents job is not therefore thought to be necessary during the valuation exercise.

Looking at the relationship between assessments of ride and those of other attributes, there may be a weak Halo affect explaining at most 4-5% of the variance of the ride assessments. Individuals' assessments of ride and some other attributes were closely related - these are the attributes one would expect to be related from previous research (British Standards (1987), International Standards Organisation (1985)). The closely related attributes appear to be (in approximate order of explanation) noise, seats, heating and ability to read. In the worst case we can say that the British Rail ride assessments have been contaminated by approximately 42%.

There are a number of reasons for believing such interference estimates are biased - probably upwards. The estimates had to be made with parametric statistics (for which the data is not suited), the comparisons used to produce these estimates are based on differences in peoples' perceptions rather than changes in engineering levels (section 3.3) and the rating methods used involved a large number of repetitions using a very general scale.

One would have expected ability to read and view to be more important. The former was widely misunderstood and view perhaps never became restricted enough to interfere with passengers' adjustments to the motion of the train. Overall these results are what one would expect from previous research.

There are two related problems with this type of repeated attribute assessment. Respondents may assess all attributes in a similar way, either because they have taken little care with their responses and thus give only an overall impression, or because they are unable to isolate attributes. However results described earlier (section 3.2.8) suggest that most respondents have taken care over their assessments.

5. CONCLUSIONS:

The research described in this chapter has provided a number of useful insights into the ride issue. The relationship between assessments of ride and those of other attributes has been considered, generating an approximate overall value for an interference effect. An indication of the sensitivity of people in detecting differences in engineering ride has been gained. Finally some form of relationship between ride assessments and personal/trip characteristics has been discovered.

These findings, which have been developed from a very large data set, provide a useful background for the analysis contained in the next chapter: where a survey designed specifically for this thesis is analysed. The results of this section, even when considered in isolation, make a useful contribution to reaching the objectives of the research described in the first chapter.

Chapter Seven Detailed Ride Investigation

1. INTRODUCTION:

In chapter three a number of possible ride valuation approaches were identified. Some of these approaches have been eliminated as a result of the preliminary enquiry contained in chapter four. The approaches that remained were then discussed in detail and it became clear that each was associated with a series of unknowns, that could represent potential weaknesses. These unknowns and methods of investigating them are discussed in chapter five. In the current chapter a series of experiments are conducted in an attempt to solve these issues. The results coupled with the information generated in chapter six (where a more general investigation is undertaken) should allow the number of possible valuation approaches to be reduced to a manageable level, by eliminating those which appear the least promising.

Much of the investigation is centered around the relationship between engineering ride and perceptions of ride quality. The study was implemented in two parts (in April and July 1989) as this allows an investigation of ride on a number of types of stock.

The Bristol to Weymouth line in the South West of England was considered the most suitable for these investigations. It was chosen as the route contained a number of different track qualities and was long enough to allow effective interviewing. This line was also having its old trains replaced with Sprinters and could be surveyed with the limited resources available. This location is considered in more detail in appendix three.

2. DESIGN OF INITIAL SURVEY:

2.1. General Questionnaire Issues:

Self-completion questionnaires were distributed and completed on-train, though some respondents completed a further section at home. All parts of each questionnaire were numbered, so that a record could be kept of when and where they were handed out. To check that respondents had the necessary experience, all were initially asked where they were travelling to/from. Only those passengers who would be (had been) travelling over the relevant sections of the route were given questionnaires.

Respondents were divided into two groups. Each group were given slightly different questionnaires, these are shown in appendix five and six. The first group completed their questionnaires while travelling over the sections being assessed (During assessments) and were therefore hyper-sensitised. The second group completed their questionnaires at the end of the sections (After assessments) to be assessed. Obviously the two types of questionnaire were not mixed on the same train, ensuring that After respondents were not alerted to the procedure by the actions of During respondents.

All questionnaires were piloted in early April 1989. The pilot was successful and only minor modifications had to be made to the questionnaires. The open question asking about the frequency of train travel was replaced by a closed question, as respondents kept giving unspecific replies (for example, occasionally or sometimes). This problem was expected, but not on such a scale. The questions asking for assessments of ride on specific sections of the route were clarified. Finally the statement, "This train" was underlined to stress that questions were not general.

After the real survey some other points were noticed, that would require modification if the questionnaire was to be used again. The quality of responses would have been improved, if an example had been put on the graphic scales. Of the initial batch of questionnaires approximately 20% were incorrectly completed. As a result an example was handwritten onto each of the final batch of questionnaires - this clearly improved the quality of the responses.

One part of the questionnaire asked respondents to assess certain track sections; this stated that if were unable to do this they should scribble the question out. Very few people did this (though one or two commented on the difficulty of the exercise). It is suspected that people were reluctant to appear unhelpful by not answering the question and so they made their best attempt. It may have been better to provide a box to tick if the task was considered too difficult.

One or two respondents thought the textual ride scale was a representation of the route and that they were supposed to mark rough sections of track. To combat this and any other misunderstandings, the task was explained to most respondents when the forms were handed out. But as attention spans could be short and there was a need to get the questionnaires out quickly (especially the During ones) this introduction had to be kept brief.

The timetable on the line made it very difficult to use resources efficiently and even after careful planning a number of hours were spent waiting for trains at various station on the route. During the week trains were lightly loaded on the southern section of the route with very few through passengers, this made it even more difficult to obtain an adequate sample.

Ideally the line should have been ridden on before the survey was designed, rather than relying on experience of past journeys and British Rail information. More piloting may also have been beneficial. There was also evidence of the line being over-surveyed, as another set of questionnaires were issued earlier in the week for a British Rail contract.

All data from the questionnaires was coded and entered into a computer for analysis. Positions on the textual ride scales were converted to values between one and nine using a specially designed stencil.

2.2. Choice of Engineering Ride Levels for Assessment:

2.2.1. Initial Approach:

Initially 1960's era stock was assessed, as the line selected for the study had not then been allocated new trains. Textual locations were derived later for the new Sprinter trains. The various types of old D.M.U. operating the line have been treated as one group. Engineering ride was assumed to be the same for all these types of stock (most of these trains used the same or similar suspensions). The use of these different 1960's trains could have increased the variability of ride ratings because of interference effects though there were few differences between trains.

After an initial investigation it was decided that the Bristol-Weymouth line could be broadly divided into two types of track quality. For any investigation of the linearity of ride valuation to be made (which would need three or more textual locations) further ride assessments would have to be produced using different rolling stock. The two types of track were: high quality sections, characterised by flat-bottomed continuously welded rail and poorer quality sections using jointed (sometimes bullhead railed) track.

So that each respondent could consider both types of track a stretch of line had to be found that included both types. We also needed a location where a high proportion of passengers would be making long distance trips and were therefore likely to have experienced all sections. Finally it was important that a section of line was chosen where the trains were not usually full (biasing the response) so that interviewing would be practical.

As a result the line between Westbury and Yeovil Pen Mill was selected. The first part of the section (Westbury to Castle Cary) was considered to have track of higher quality than the second part (from Castle Cary to Yeovil Pen Mill).

As some respondents would be asked to assess ride while travelling these sections it was important that the sections were not too long. Travel between Westbury and Yeovil Pen Mill took forty-five minutes, it was thought that respondents would be able to concentrate for this period.

Some respondents may have known that the Westbury - Castle Cary section was mainline (and the Castle Cary - Yeovil Pen Mill section was not) this knowledge may have influenced their perceptions of ride. It has to be assumed that most people had insufficient railway knowledge for this to happen. The exercise also assumes that each respondent sat in the same position on the train, for the whole of his journey.

Ride was assessed on southbound trains only, so that passengers' ride experiences would be comparable. Southbound passengers would have come from a greater variety of origins, allowing the effects of trip length to be investigated. Covering southbound trains on the southern section of line, meant that all questionnaires would be retrieved, as the train terminated with the interviewer on-board at Weymouth. The interviewer could stay with the train and follow it north ready to start the next set of interviews: so no time was wasted waiting for another train.

It was also decided to avoid northbound assessments, as the very southern section of the route involves the train ascending a very steep gradient over good track: the excessive noise and vibration over this section could have interfered with ride assessments.

2.2.2. Findings From The Survey:

Although the chosen sections of line had track of different standards, the whole of the route was originally built as a main line. All sections therefore had similar curvature throughout their length, making the difference in engineering ride between sections small.

Jointed track did appear to make a significant difference to engineering ride measures. But it did not occur in sufficient concentrations to make the ride differences between sections as great as desired. In fact the interviewer could not really tell the difference between overall sections of track and some respondents commented on their inability to establish differences between sections of route.

With hindsight it may have been better to have considered a shorter section, to compare with Castle Cary to Yeovil Pen Mill. The first part of the Westbury to Castle Cary section (Westbury to Frome) exhibited quite poor ride and could have been omitted, giving a greater contrast with the poorer Castle Cary - Yeovil Pen Mill section. As the first section had two intermediate stops and the second none, this could have affected the assessments.

Considering shorter sections of route would have meant that more passengers had appropriate ride experience. It turned out that on weekdays very few passengers made through journeys, so obtaining a sample of travellers with sufficient experience to answer the questionnaires proved difficult. Generally on weekdays only three or four respondents were obtained on each train, though this rose to thirty on weekend trains.

It took approximately thirty minutes to travel from Westbury to Castle Cary; this may have been too long to retain the interest of respondents producing assessments while they experienced the section. A few respondents actually forgot about the exercise during this time.

If a shorter section had been selected it would have started at either Frome or Bruton. Neither of these stations are as important, or likely to be as easily remembered by respondents, as Westbury which is an interchange and big junction. It would also have been difficult to get all the During questionnaires out between Bruton and Castle Cary. Overall it would probably have been better to use Frome - Castle Cary (20 minutes) as the first section.

2.3. During Questionnaires:

The During questionnaires were given out (southbound) as the train left Westbury and collected after Yeovil Pen Mill. As these questionnaires were given out during the first part of the route to be assessed, each respondent (on any train) would start the task at a slightly different time. Thus each respondent would be assessing slightly different sections of track. This also means that the After and During respondents are not strictly assessing the same track sections, as the former were done after respondents had ridden the whole section concerned.

Yeovil Pen Mill turned out to be the most effective place to get off southbound trains and switch to northbound trains for the next run. To make the most productive use of time, During questionnaires were collected as the train drew into Yeovil Pen Mill. This meant that it would have been better if the During questionnaire had the ride assessment as the very last question, so that they could be collected quicker under these new conditions.

Distributing all the questionnaires on a run rarely took more than ten minutes, collecting questionnaires was considerably quicker. This and the small ride differences within each section mean that it is assumed, there is no significant difference in respondents' ride assessments, as a result of questionnaire distribution and collection times.

A few respondents were observed filling in the During questionnaire too early, which would obviously make the results meaningless. The During survey really required too much effort and contained too many unreliabilities to be used with unpaid respondents.

2.4. After Questionnaires:

With the After group the interviewer boarded southbound trains at Yeovil Pen Mill. The questionnaires were distributed between Yeovil Pen Mill and Dorchester West, so that none of the assessments were hyper-sensitised. The fact that the interviewer had not have been travelling throughout the route, should have also helped to contain hyper-sensitivity. There was fifty minutes before the train reached Weymouth, for the questionnaires to be completed.

Overall ride perceptions are likely to be significantly affected by the most recent experiences. To try and ensure recent experiences were controlled over all respondents, questionnaires were only issued between Yeovil and Dorchester West (which had similar track to the last section being assessed).

The After respondents produced two types of ride assessments. They assessed ride over the two sections and also gave an overall impression of ride up to the point they were given the questionnaire. For assessments to be comparable all questionnaires should have been completed at exactly the same time, this was clearly not possible. However the period during which questionnaires were issued was small compared to the duration of overall trips and so any effect is likely to be small.

The After group were also given a take-home question, this asked them to assess the importance of ride off-train (appendix seven) and allowed a comparison with the same question asked on-train - allowing the importance of environmental effects to be estimated.

The take-home section was responsible for most of the problems with the After group. Some people looked at this section while on the train and some filled it in on-train and handed it back to the interviewer. These questionnaires had to be destroyed, as they no longer tested the effect of the decision environment. It may be that some people completed the take-home section on-train and then returned it by post. The extent of the resulting error will never be known, but it is thought to be small.

Passengers were instructed verbally, as well as on the take-home section to do the exercise at home and it is considered likely that someone who obeys the instruction to post that section, will also obey the instruction to complete it at home. Further, many passengers were observed putting the take-home section away at the beginning of the exercise. It was clear that those who did read the take-home questionnaire on-train, were confused as to why they would be answering the same question again. Even so the postal response rate was promising (54%).

2.5. Sampling:

Samples were taken on a number of trains, representing peak and off-peak travel. A random sample of passengers was to be taken, but so few passengers were eligible for the survey that this proved unnecessary. The desired sample sizes could not be established during the design of the survey, as the variation of individuals' perceptions (for a given level of engineering ride) was unknown. As a result the biggest samples possible with the available resources, was obtained. Most passengers were happy to complete a questionnaire, only approximately 5-10% did refuse.

The full scale survey was carried out over two Saturdays (three trains in total), one Sunday (one train), a Monday (four trains), a Tuesday (four trains) and a Wednesday (four trains). All trains were covered during the survey except the 06.29 and 22.01 arrivals at Yeovil Pen Mill. According to staff on the line both these trains ran virtually empty down to Yeovil Pen Mill. The 6.29 was also rostered to a class 150/2 Sprinter during the period of the survey and so was not appropriate to this stage of the survey. For the second set of ride assessments, on replacement trains, attempts were made to cover trains at the same or similar times.

TABLE 7.1: QUESTIONNAIRES GIVEN OUT AND RESPONSE RATES.

TYPE	NUMBER ISSUED	NUMBER RETURNED	% RETURN
During	55	53	96%
After	119	119	100%
After - Postal	119	64	54%

It is believed that a reasonable representation of the longer distance passengers on the line has been obtained. It may be that shorter distance travellers have a different perception of ride quality, some indication of this should come from the variety of trip lengths considered in the data. Further the valuation of ride will use a better representation of the public and as long as perceptions are consistent this should not cause a problem.

3. INITIAL ENGINEERING RIDE MEASUREMENTS:

3.1. General Approach:

Engineering ride measurements were taken using a Macmeter (passenger ride comfort meter) borrowed from the Railway Technical Centre, Derby. This equipment provided ride measures for each section of track desired. All engineering measures are stated as I.S.O. weighted m/s2 R.M.S, as described in chapter two. Two runs were made with the Macmeter, both were taken on the same coach and in the same position within the coach. Measurements were taken between every pair of stops on the route.

Measuring began as soon as the guard sounded the buzzer, authorising the driver to pull away from each station. Measuring stopped and readings were taken immediately the train stopped at a station. These station to station figures can be aggregated to provide engineering ride measures for a longer stretch of the route. Such aggregation may give a slight bias, as an overall ride figure should include the ride while halted at stations. However, as all measurements were taken consistently and later measures were taken in the same way, all measures are comparable. Measures from sections of route were amalgamated by taking a weighted average of each section concerned. The average could be weighted by either distance, or time. Time has been selected as this indicates the duration of the experience the passenger has been exposed to, distance would not do this.

From the two sets of ride measurements, an average figure has to be obtained for each section of route, which is the best representation of engineering ride on that section at the time of the interviews. If any section, on either run, was affected by abnormal signal checks etc. that figure was not used in the calculations. For all other sections the mean was taken.

3.2. Limitations:

It was important that measurements were made of every section of the route and to ensure equipment was not tampered with. Engineering measures of ride could therefore not be produced on the trains on which passengers were interviewed. There are a number of factors that could mean the engineering measures are unrepresentative of the levels felt by respondents.

Some intermediate stations were not included in the measurements, these were: Freshford, Avoncliff, Thornford, Yetminster and Chetnole. These stations were not always called at during the research and none of the through passengers, we are concerned with, boarded at these stops. Where one of these stops was encountered the Macmeter was kept running throughout the stop.

These stops and the associated periods of slowing and accelerating away, would reduce the severity of the ride figures over some sections. But, the time the train spent at these stops was small and so would not have had too dramatic an effect on the results. This effect can be seen between Bath Spa and Bradford on Avon, on the first run there were no intermediate stops (overall reading: 15 and 40), while on the second there were two (overall reading: 12 and 39). The second measurements are slightly lower as a result. The figures used in the analysis will have been affected less, as the two main sections considered (Westbury to Castle Cary and Castle Cary to Yeovil Pen Mill) did not cover any excluded stops. The other ride sections considered would be longer than Bath Spa to Bradford on Avon, diminishing the effects of excluded stops.

During the setting up of this exercise, it was learnt from British Rail Derby that ride characteristics may be different between powered and unpowered coaches, of first generation D.M.U.s and within each coach. Ride was expected to be most lively in powered coaches because of the stiffer suspension and engine vibration. Ride was also expected to be rougher at the uncoupled ends of the coaches. Different classes of D.M.U. of the same era were not expected to display major differences in ride, though maintenance could have an effect.

The above factors meant that: the class of the unit, the seating position and whether the unit was powered or unpowered, were noted on each questionnaire. As the rolling stock on the line was in poor condition, units were often made up of coaches of different classes and so the class of each coach had to be noted. The seating position was placed into one of three categories: middle, loose-end (not coupled to another coach) or tight-end (coupled to another coach). Most of the ride assessments and all engineering ride measurements were done with powered cars, which were predominant on the line because of severe gradients.

It was confirmed during the ride measurements, that the passenger loading on each coach could also affect ride. This loading effect is apparent between Bristol Temple Meads and Bath Spa (table 7.2). The engineering ride measures may not be strictly representative, as most were taken on heavily loaded trains. Though most of the interviews took place in similar circumstances.

Track maintenance over the weekend of 22/23rd April 1989 improved the quality of the Westbury to Castle Cary section (new rails and ballast were installed). Forty After questionnaires were done before this improvement in ride occurred.

For the analysis to be continued it has to be assumed, the engineering measurements made are representative of the trains that passengers were interviewed on.

TABLE 7.2: ENGINEERING LEVELS OF RIDE FOR WHOLE ROUTE.

Most representative engineering ride values, derived from appendix nine. Also the mean time taken over each section from the British Rail timetable.

SECTION	LATERAL	VERTICAL	MINUTES
Bristol Temple Meads - Keynsham	15.5	38.5	7
Keynsham - Oldfield Park	12.5	37.5	8
Oldfield Park - Bath Spa	10	33	4.5
Bath Spa - Bradford On Avon	12	39	15.5
Bradford On Avon - Trowbridge	10	35	6.5
Trowbridge - Westbury	9	32.5	7
Westbury - Frome	14	36.5	10
Frome - Bruton	15	41.5	13
Bruton - Castle Cary	10.5	32	6
Castle Cary - Yeovil Pen Mill	23	40	15
Yeovil Pen Mill - Maiden Newton	14.5	37	23.5
Maiden Newton - Dorchester West	18.5	41.5	10.5
Dorchester West - Upwey	12	39.5	9
Upwey - Weymouth	11	34	4

TABLE 7.3: ENGINEERING LEVELS OF RIDE ON A MODERN TRAIN

For comparison, three readings taken on Class 442, "Wessex Electrics" over the same route (northbound).

SECTION	LATERAL	VERTICAL
Weymouth - Upwey	5	9
Upwey - Dorchester South	5	9
Dorchester South - Moreton (at speed)	8	12

TABLE 7.4: ENGINEERING LEVELS OF RIDE DERIVED FOR SECTIONS OF INTEREST.

SECTION	LATERAL	VERTICAL
Westbury - Castle Cary	13.7	37.8
Castle Cary - Yeovil Pen Mill	23.0	40.0
Bristol Temple Meads - Yeovil Pen Mill	14.2	37.6
Keynsham - Yeovil Pen Mill	14.1	37.5
Oldfield Park - Yeovil Pen Mill	14.2	37.5
Bath Spa - Yeovil Pen Mill	14.5	37.8
Bradford On Avon - Yeovil Pen Mill	15.1	37.4
Trowbridge - Yeovil Pen Mill	15.8	37.7
Westbury - Yeovil Pen Mill	16.9	38.6
Frome - Yeovil Pen Mill	17.7	39.2
Bruton - Yeovil Pen Mill	19.4	37.7

Looking at the two sections compared by respondents there is a clear difference in the Macmeter readings, especially laterally. Comparing individuals' perceptions of ride for overall journeys down to Yeovil Pen Mill was unlikely to produce any results, as there are only small differences in engineering ride over such long sections.

4. RESULTS OF INITIAL SURVEY:

4.1. Effects of Personal and Other Characteristics:

Previous research (outlined in chapter two) has identified a series of factors that may influence individuals' assessments of ride quality. It is important to gauge the extent of such relationships, in the present context, as they could have an important effect on any findings produced during this research. Cross-tabulations are therefore presented between respondents' ride assessments and other variables identified in previous research.

A number of ride assessments were produced by the research and so a choice had to be made of which one to use in the cross-tabulations. The chosen assessment was the After assessment of the Castle Cary to Yeovil Pen Mill section. This was selected as more assessments of this kind were made and because the variation of individuals' assessments seemed to be no greater than any other assessment - suggesting that these ride assessments were as reliable as any other (section 4.2.1). Of one hundred and nineteen After respondents ninety produced valid ride assessments.

Two statistics were used to test the relationship between ride and the other variables. These were the Chi-Squared test and Kendall's Tau-C correlation coefficient (where applicable). To ensure that the Chi-Squared results were valid it proved necessary in some cases to merge groups. For all these comparisons a null hypothesis was constructed, "There is no significant relationship between the ride assessment and the other variable".

4.1.1. Ride By Type Of Unit:

CHI-SQUARE INVALID.

An attempt was made to see if there was a relationship between powered and unpowered cars and the respondents assessment of ride quality. Unfortunately the data set proved too small for the test to give a valid result (Groebner and Shannon (1985)). The minimum expected frequency being less than one and 50% of the cells had an expected frequency below five. Most of the coaches used on the line were powered and so few respondents were interviewed on unpowered cars.

4.1.2. Ride By Position:

CHI-SQUARE INVALID.

The result of this test again proved invalid, but only just, the minimum expected frequency being 0.978 and 20% of the cells having expected values below five. Some credence may therefore be attached to this result (which was 91% significant). This result suggests that assessments made by respondents at the end of the coach are more severe than those made by respondents in the middle of a coach. This is the result that would be expected, as engineering ride is worse at the end of a coach.

4.1.3. Ride By Experience Of Other Lines:

CHI-SQUARED INVALID.
TAU-C 86% SIGNIFICANT.

The result of the Chi-Squared test again proved invalid, but only just, the minimum expected frequency being 0.976 and 20% of the cells having expected values below 5. This result (which was 91% significant) may therefore have some value. The correlation proved valid (and significant) being 0.135. This suggests that people with less experience of rail, think ride is smoother.

4.1.4. Ride By Personal Characteristics:

A number of cross-tabulations were performed between assessments of ride quality and personal/trip characteristics, those considered were: age, sex, employment, regularity of rail use and trip purpose. Unfortunately the small size of the data set and that fact that many respondents were in the same category, meant that few results were valid. Only the age comparison gave any useful results (Tau-C 83% significant): this suggested that people in the 16-24 age group are considerably more critical of ride. This result is remarkably similar to that produced by the previous analysis in chapter six.

4.1.5. Ride By Need To Stand On The Current Trip:

CHI-SQUARED INVALID.
TAU-C NOT SIGNIFICANT @ 80%.

The low significance of the result and the small sample of standees means that no conclusions can be drawn here.

4.1.6. Ride By Experience Of Standing Generally:

CHI-SQUARED INVALID.
TAU-C 98% SIGNIFICANT.

This is a very significant relationship, suggesting that passengers with more experience of standing are more critical of ride quality. As standees experience greater accelerations we would expect this result.

4.1.7. Ride By Importance Of A Ride Improvement (On Train):

CHI-SQUARED INVALID.
TAU-C 99.1% SIGNIFICANT.

This is a very strong relationship. A strong negative correlation of -0.202 was recorded. This means that respondents who thought the current ride was smoother were less likely to desire an improvement. Although this seems obvious it does at least give an indication of respondents being consistent, increasing the validity of other results.

4.1.8. Ride By Importance Of A Ride Improvement (At Home):

CHI-SQUARED INVALID.
TAU-C NOT SIGNIFICANT @ 80%.

This relationship is not strong enough to produce any significant conclusions. This result is considerably weaker than the on-train relationship (section 4.1.7) suggesting a possible change in attitude between the first response and that recorded when the passenger reached home. But not all the respondents will have filled in the take-home questionnaire the same time after the journey. Some questionnaires may have been done on-train or on the way home, others may have been done several days after the event.

4.2. Effects of Hyper-Sensitivity:

It was important to discover to what extent passengers being hyper-sensitised would affect their assessment of ride. It is also useful to know which form of assessment gives the most reliable results.

These aims were achieved by testing the difference between the variance of ride assessments with the F test, to see which form of measurement is associated with the least spread (thus which is the most reliable).

The difference between mean ride scores was originally to be tested using a t test, but the data was far from normally distributed (see figure 7.1) making a t test invalid. As a result, the (non-parametric) Mann-Whitney U test was used to distinguish between the two independent groups.

4.2.1. Difference Between Variances:

First a null hypothesis was created, "There is no significant difference between the variances of the After and During ride assessments".

<u>Castle Cary to Yeovil Pen Mill:</u>

During Standard Deviation = 1.91   Variance = 3.65
n = 52
After Standard Deviation = 1.99   Variance = 3.96
n = 90

F RATIO = 1.09
D.f. Numerator = 89   D.f. Denominator = 51

Not Significant @ 90%.

Westbury to Castle Cary:

During Standard Deviation = 1.95   Variance = 3.80
n = 52
After Standard Deviation = 2.30   Variance = 5.29
n = 90

F RATIO = 1.39
D.f. Numerator = 89   D.f. Denominator = 51

Not Significant @ 90% but close.

The During assessments would be expected to be more reliable, as passengers were aware what they were assessing throughout the journey. But there is no significant difference between the variation of the During assessments and those of the After assessments at a 90% level of confidence. The null hypothesis cannot therefore be rejected.

An F test was also done to see if there was a difference in the reliability of the first and second After assessments. No difference would be expected between the two During assessments as both were made under the same conditions. But with the After approach the first assessment involves a longer period of recall and thus potentially greater scope for error. A null hypothesis was created, "There is no significant difference between the variances of the first and second, After ride assessments".

Castle Cary to Yeovil Pen Mill verses
Westbury to Castle Cary:

Castle Cary to Yeovil Pen Mill (After)
Standard Deviation = 1.99   Variance = 3.96   n = 90

Westbury to Castle Cary (After)
Standard Deviation = 2.30   Variance = 5.29   n = 90

F RATIO = 1.36
D.f. Numerator = 89   D.f. Denominator = 89

Not Significant @ 90% but close.

Again the null hypothesis could not be rejected at a 90% level of confidence.

Two of the above F statistics (1.39 and 1.36) are close to the 90% confidence limits (1.48 and 1.39 respectively). We can thus say that there is some indication that the After assessments of the first section experienced by respondents, are less reliable than the After assessments of the last section experienced. We can also say that the After assessment of the first section experienced by respondents, is less reliable than both During assessments. This difference is probably a result of increased recall error, associated with the assessment of the first section experienced.

This result has important implications for the design of the rest of the study. The After assessments are likely to be closer to the decision process, as passengers will make their decision to travel after and not during a journey. As it seems possible to get results with an After assessment that are as reliable as a During assessment, the most realistic assessments can be used to generate ride values.

4.2.2 Difference Between Means:

A null hypothesis was constructed, "There is no significant difference between the During and After sets of scores".

<u>Castle Cary - Yeovil Pen Mill:</u>

During Mean = 3.29   n = 90, After Mean = 4.06   n = 52

Mann-Whitney value = 2138.5   M.W. Mean = 2626
Standard Deviation (Corrected for ties) = 256.29
Deviation from Mean = 487.5 1.90 of a Standard Deviation.

Significant @ 94%.

<u>Westbury - Castle Cary:</u>

During Mean = 3.60   n = 90, After Mean = 4.36   n = 52

Mann-Whitney value = 1854   M.W. Mean = 2385
Standard Deviation (Corrected for ties) = 256.29
Deviation from Mean = 531   2.26 of a Standard Deviation.

Significant @ 98%.

It is clear from the Mann-Whitney test that the null hypothesis can be rejected. There is a very significant difference between the two sets of scores. During assessments are more critical of ride (lower values) than After assessments done from memory. The Castle Cary to Yeovil Pen Mill During mean is 23% greater than the After mean. The Westbury to Castle Cary During mean is, similarly 21% greater than the After mean. These more critical During results are consistent with previous research.

4.3. Sensitivity of Ride Perceptions:

Each respondent was asked to assess two adjoining sections of track. Engineering ride measurements showed that there was a difference in ride quality between the sections. It is therefore important to find whether respondents were able to perceive this difference correctly. As the distributions appeared not to be normal, a (non-parametric) paired Wilcoxon test was used to test for significance. A null hypothesis was generated, "There is no significant difference between the respondents' assessments of the two track sections".

<u>After Assessments:</u>

Westbury to Castle Cary: Mean = 4.36
Castle Cary to Yeovil Pen Mill: Mean = 4.06

Wilcoxon value = 333   Wilcoxon Mean = 473   n = 90 pairs
Standard Deviation (Corrected for ties) = 81.35
Deviation from Mean = 140   1.72 Standard Deviation.

Significant @ 91%.

<u>During Assessments:</u>

Westbury - Castle Cary: Mean = 3.60
Castle Cary - Yeovil Pen Mill: Mean = 3.29

Wilcoxon value = 180.5 Wilcoxon Mean = 232.5
n = 52 pairs
Standard Deviation (Corrected for ties) = 47.24
Deviation from Mean = 52 1.10 of a Standard Deviation.

Not Significant @ 80%.

It is clear in both cases that Castle Cary to Yeovil Pen Mill was assessed most critically by respondents, this corresponds with the engineering measurements of ride. However the null hypothesis could only be rejected with the After approach. The non-significant During result could be a result of the smaller sample size or the perceptions of one section interfering with those of the other section.

4.4. Consistency of Ride Perceptions:

We wanted to find whether respondents ride perceptions were consistent - whether a respondent who perceived one section to be rougher than the mean, would perceive other sections to be rougher than the mean and vice-versa. Spearman's (non-parametric) rank correlation was used to establish the consistency of individuals estimates. The assessments of both sections of route were correlated, for After and During groups. A null hypothesis was generated, "There is no significant relationship between the first and second assessments, produced by each respondent".

<u>During:</u>

Spearman's Rank Correlation = 0.77

Significant @ 99.99%.

<u>After:</u>

Spearman's Rank Correlation = 0.63

Significant @ 99.99%.

The null hypothesis was easily rejected. It is clear from these very significant positive correlations that people are consistent in their assessments of ride quality. Someone who is more critical of one section of track will be more critical of any other section of track. This close relationship also infers that the results are valid, as people would be expected to be consistent in their perceptions.

4.5. Effect of the Decision Environment:

We wanted to establish whether the environment in which a travel decision was made, would affect the outcome of the travel decision. It may be that ride is assessed more or less critically when the respondent is off the train.

To investigate these effects respondents were first asked where they made their decision to travel today. The response would indicate the correct environment in which to replicate any travel decisions. The results showed clearly that most travel decisions were made at home, followed by work.

TABLE 7.5: WHERE DECISION TO TRAVEL WAS MADE.

LOCATION	FREQUENCY	%
Home	32	63
Work	9	9
Station	0	0
Street	0	0
Other	10	10

Each respondents was next asked to assess the importance of an improvement in ride quality, both on-train and at-home. Any difference in ratings would indicate the effect of the decision environment. As the distributions appeared not to be normal a (non-parametric) paired Wilcoxon test was used to test for significance. A null hypothesis was constructed that, "There is no significant difference between the scores at home and on the train".

At Home: Mean = 6.46
On-Train: Mean = 6.17

Wilcoxon value = 186 Wilcoxon Mean = 232.5
n = 52 pairs
Standard Deviation (Corrected for ties) = 47.68
Deviation from Mean = 52 0.98 of a Standard Deviation.

Not Significant @ 80%.

The null hypothesis could not be rejected and so we cannot say that the different environments had an effect on respondents' assessments of the importance of ride.

It interesting to note the response to the ride improvement question for its own sake. The values presented are the on-train results.

TABLE 7.6: THE IMPORTANCE OF A RIDE IMPROVEMENT.

IMPORTANCE		FREQUENCY	%
Very Important	9	27	23
	8	16	13
	7	16	13
	6	7	6
	5	11	9
	4	7	6
	3	6	5
	2	7	6
Not Very Important	1	15	13

There is a clear desire for improvement on the ride quality provided by the old D.M.U.s then operating on the line.

4.6. The effect of Trip Duration on Ride Assessments:

Respondents were asked where they had got on the train so that an estimate of the duration of their trip (up to the point where they were interviewed) could be made. To see whether the duration of the trip made a difference, this was correlated with the two After ride assessments. Spearman's (non-parametric) rank correlation was used. A null hypothesis was generated, "There is no significant relationship between respondents' ride assessments and the duration of their trip".

Westbury - Castle Cary: $R = 0.06$
Not Significant @ 80%.

Castle Cary - Yeovil Pen Mill: $R = 0.03$
Not Significant @ 80%.

The null hypothesis could not be rejected.

4.7. Experience of other lines:

We wanted to know how much experience respondents had of other local lines. This would give an indication of whether it would be possible to compare ride between different lines, in a ride valuation exercise.

TABLE 7.7: EXPERIENCE OF OTHER LINES.

RESPONSE	NUMBER	%
Yes	81	50.3
No	80	49.7

Approximately half the sample had experience of other lines. But some of them may still have had insufficient experience to recall ride quality on the other lines.

4.8. Comparisons of Overall Ride assessments:

All After respondents provided an overall ride assessment for the whole of their journey up to Yeovil Pen Mill. Respondents were asked where they boarded the train, so that people who had travelled over the same sections of line could have their assessments grouped. Only some of the overall assessments could be used in the analysis, as the sample sizes were too small in many cases. The only three origins with sufficient numbers of passengers were: Bristol, Bath and Westbury.

It was thought that asking respondents to do an overall appraisal, would produce more reliable assessments of ride. But it is clear from comparing the standard deviations in the table below with those for the set sections of route (1.95 and 2.30) that this is not case.

TABLE 7.8: RIDE ASSESSMENTS FOR WHOLE JOURNEY.

ORIGIN	MEAN	STANDARD DEVIATION	n
Bristol Temple Meads	3.71	2.27	45
Bath Spa	3.36	1.83	25
Westbury	4.19	2.42	27

To see if there was a significant difference between the three assessments, a (non-parametric) Kruskall-Wallis One-Way Analysis of Variance (ANOVA) was used. This was chosen rather than a series of Mann-Whitney U tests, to keep the possibility of a type I error low (Hayes (1988)). A null hypothesis was generated, "There is no significant difference between the scores of the three groups".

Kruskall-Wallis Statistic (corrected for ties) = 1.37

Not Significant @ 80% with 2 d.f.

The null hypothesis was not rejected. Overall no significant differences emerge. This is not surprising when one considers the small differences in the overall journeys considered (table 7.3). This result coupled with assessments of the difference between track sections, does give us some indication of the threshold of ride perceptions. It is clear that if different track sections are to be compared, large differences in engineering ride must be apparent between sections for respondents to notice. Even then the results may not be as clear as desired.

5. SECOND SURVEY - ASSESSMENTS OF ENGINEERING RIDE FOR OTHER ROLLING STOCK:

5.1. Introduction:

The next stage of the research was to get passengers to assess other types of rolling stock. This would allow us to see if the ride differences between trains, were more perceptible than between sections of track. This later set of data would also allow an investigation into the problems of isolating ride from other attributes.

In May 1989 most of the stock on the Bristol to Weymouth line was replaced. The replacement stock was a mixture of class 150/2 and 155 Sprinters, locomotive hauled Mark 2 abc and Mark 2 def coaches, a 4TC unit and some old D.M.U.s.

TABLE 7.9: ROLLING STOCK ASSESSED AND ABBREVIATIONS
(Cooper (1984), Haresnape (1985), Haresnape (1986)).

NAME	DATE BUILT	DESCRIPTION
Old D.M.U.	1956-60	Old style Diesel Multiple Unit.
Mk1 RMB	1951-63	Buffet car.
Mk2 abc	1966-71	Open coach with sliding windows.
Mk2 def	1971-76	Open coach with air conditioning.
150/2	1986-88	Original Sprinter.
155	1987-89	Enhanced Sprinter.

The Mark 2 coaches were either air conditioned (def) or conventionally ventilated with opening windows (abc). All Mark 2 coaches use the same underframe and bogies and are thus characterised by the same levels of engineering ride. Both types of Sprinter use the same bogies and are also characterised by very similar levels of engineering ride. Overall the Mark 2 coaches should be slightly smoother than a Sprinter. The 4TC unit uses the same bogies as the Mark 2 coaches and so should share their ride characteristics. Some difference would be expected between vehicles of the same type due to maintenance, but on average these generalisations are correct. These engineering ride measurements are reported in a Railway Technical Centre report (Frederick (October 1987)).

5.2. Survey Design:

It was thought that some interesting comparisons could be made between the various types of stock. The two types of Mark 2 coach would allow the examination of the effect of noise on ride perceptions, as the air conditioned coaches are considerably quieter - having no opening windows. The two types of Sprinter have markedly different interiors and this may give some indication of interference effects.

The operating diagrams for the various stock types were obtained, so that a survey pattern could be established. It became clear that not all stock types could be covered. This was because some of the stock ran rarely and/or at awkward times. It was also important to try to maximise sample sizes. For these reasons it was decided to concentrate on, both classes of Sprinters and both types of Mark 2 coach. As the new trains had been in service for some time, any newness effect should have worn off.

For this survey, all that was required was the generation of ride perceptions, that were comparable with those produced in the previous survey. Data was therefore only required for one section of line. This was chosen as Castle Cary to Yeovil Pen Mill, as this was the section for which the most reliable ride measurements were obtained on the old D.M.U. Stock. The After, rather than the During approach, was used as this was felt to be more representative of the decision process. In the previous survey generating only one assessment from memory, proved just as reliable as the production of a single During assessment. For the second survey passengers were thus asked for their opinion, at the end of the appropriate track section (that is south of Yeovil Pen Mill).

Only assessing one section made sampling considerably easier than with the previous survey, as respondents did not need to have covered such a long stretch of route to be eligible for the exercise. The later questionnaire was much more straightforward as less information was required and much experience had been gained with the implementation of the previous survey. The questionnaire comprised of: a note of the unit on which the interview took place, a space for general comments about the service on the line and the ride assessment (appendix eight).

The ride assessment question was improved in the later survey and a clear example was incorporated into the question. This new form of question was tested on a number of students at Cranfield and was found to be successful. The new form of question was also found to be more effective in the full survey. The variation of the ride assessments was found to be considerably smaller with the new questionnaire (table 7.10). The only comparison that was not significant, was between the old D.M.U. and the Mark 1 RMB, this lack of significance is quite likely to be a result of the small sample size of the Mark 1 RMB.

TABLE 7.10: VARIANCES FOR BOTH SURVEYS.

COMPARISON (1 v 2)	VARIANCE (1)	n (1)	VARIANCE (2)	n (2)	F-RATIO	SIG.
Old D.M.U. v Mk1 RMB	3.96	90	3.57	11	1.11	N/S
Old D.M.U. v Mk2 def	3.96	90	0.24	44	16.50	99%
Old D.M.U. v Mk2 abc	3.96	90	0.61	26	6.49	99%
Old D.M.U. v 150/2	3.96	90	0.10	64	39.60	99%
Old D.M.U. v 155	3.96	90	0.23	43	17.22	99%

5.3. Implementing the Survey:

It was decided in both surveys that passengers would not be interviewed, if they were more than 10 minutes late. This decision is based on the Cranfield reliability study where a ten minute delay represented a significant jump in aggravation (Benwell (1985)). In both surveys trains were often delayed, setting off a cascade of delays and cancellations (because the line is single track). This meant that passengers could not be interviewed on some of the trains.

During the second surveys stock was often replaced at short notice and old D.M.U.s (which were not to be sampled) substituted. An overtime ban by A.S.L.E.F. also meant that some trains did not run with the expected stock. This also made it very difficult to travel to the survey area and considerable time was wasted waiting for cancelled trains. Finally the one day rail strike by the N.U.R, caused further difficulties for the research.

Some trains were replaced by old D.M.U.s with only one powered car (steep gradients mean that two were usually used) these trains got progressively later and disturbed the whole timetable. If the previous train had been cancelled interviews did not take place, as some very late passengers would be on-board and it would be impossible to distinguish these passengers from the on-time ones. Unfortunately the final attempt to get the quota on the Mark 2 abc stock had to be abandoned as this train was 20 minutes late and the previous one had been cancelled. As a result some of the sample sizes (particularly the class 155 and Mark 2 abc) are not as large as was desired.

Although the data was collected during the industrial action, most services on the Bristol to Weymouth line ran normally (trains that did not were ignored). It is possible that some passengers were interviewed on normally running trains, who had been delayed earlier that day/week and so would be biased against the service. In an attempt to counteract such biases, a general open question was placed at the start of the questionnaire. This open question (it was hoped) would allow respondents to get any grievances off their chest, before attempting the important ride assessment. Although this seems unfair, as it was not done with the previous survey, the latter survey did have a number of questions before the ride question that would have had the same effect. It should be noted that there were few comments about the rail strike in this open section.

The Mark 2 def stock was in generally better condition than the Mark 2 abc stock, the latter was dirty and had luggage racks missing and armrests loose. A number of the guards commented on the state of this stock, but it remained like this throughout the week. Part of the Mark 2 def rake contained a buffet car in which people were interviewed, this was actually a Mark 1 coach on commonwealth bogies that are approximately equivalent to the B4 bogies on the rest of the rake - though again with a markedly different interior.

To see if the samples are comparable, between this survey and the previous one, an attempt was made by the interviewer to establish the sex and age of the respondents. There may be some error here, as the characteristics were noted by the interviewer rather than being put on the questionnaire by the respondents. However it was considered important to keep the questionnaire down to one side of card and the age categories are so broad that there is likely to be little error.

The later survey will probably contain a greater proportion of leisure trips as it was nearer summer. The schools had not broken up, so any differences should not be that great. The more frequent summer timetable was operating during the second survey and this may have also affected the mix of passengers. Samples on some types of train may be biased towards certain groups, as some stock only ran at certain times of the day. For example, the Mark 2 def rake only ran on Saturday mornings.

5.4. Engineering Ride Measurements:

Macmeter readings were to taken on all types of rolling stock, it was intended to get at least two sets of readings for each type of stock. The ride measurements were (in all cases) taken just over the coupled bogey, to be compatible with the measurements taken with the old D.M.U.s in the previous survey.

It is possible that the quality of the track could have changed between the surveys, meaning that not all of the difference in engineering ride was due to the change in stock. Even so, as engineering ride measurements were taken during both surveys, perceptions can still be directly related to engineering measures.

Unfortunately it did not prove possible to get all the ride readings desired, as the Macmeter failed after giving four sets of readings. It was also proving very difficult to catch the desired trains, as the various types of Sprinter were sent down at random. It was not therefore possible to tell whether a 150/2, or a 155 would be coming down the line, making it very difficult to plan ahead.

TABLE 7.11: ENGINEERING MEASURES OF RIDE FOR CASTLE CARY TO YEOVIL PEN MILL.

STOCK	LATERAL	VERTICAL	NOTES
150/2	5	11	3/4 Empty
150/2	10	16	3/4 Empty/Seemed Faster
150/2	9	18	3/4 Empty
Mark 2 abc	7	18	3/4 Empty

TABLE 7.12: MEAN ENGINEERING LEVELS OF RIDE DERIVED FOR CASTLE CARY TO YEOVIL PEN MILL.

STOCK	LATERAL	VERTICAL	NOTES
150/2	8	15	
Mark 2 abc	7	18	Only one reading
Old D.M.U.	23	40	From table 7.3

5.5. Characteristics of Respondents:

The sex and age of most respondents was available for both surveys. These are presented below for the main groups of stock.

TABLE 7.13: SEX AND AGE OF RESPONDENTS.

	Old D.M.U.*	HAULED COACHES	SPRINTERS
AGE	%	%	%

<16	8	14	15
16-24	40	23	13
25-44	24	30	49
45-59	16	14	11
60-64	3	12	5
>65	0	7	7
SEX	%	%	%
Male	48	58	61
Female	52	42	39

* After results for compatibility.

It is clear from the table above that the Sprinter (150/2 and 155) and hauled coach (Mark 2 abc, Mark 2 def and Mark 1 RMB) groups, have smaller proportions of respondents in the critical 16-24 (section 4.1.4) age group than the old D.M.U. group. The Sprinter and hauled coach group also have a greater proportion of people aged 60 and above, who tend to be more lenient in their assessments. The differing proportions of the sexes would not be expected to have much effect on the results (section 4.1.5).

5.6. Comparisons of Ride Perceptions between Trains:

From the previous discussion it is clear that we have two groups of trains, each with different levels of engineering ride. The old D.M.U. having an inferior ride to all the other stock. The mean ride scores for each train are shown in the table below, in order of smoothness. It is clear that the old D.M.U. has the worst score. There are also differences between the other types of stock, which are all supposed to have similar engineering ride levels. It is important to test whether these differences are significant.

FIGURE 7.1: PASSENGERS ASSESSMENTS OF RIDE QUALITY CASTLE CARY - YEOVIL PEN MILL (1989).

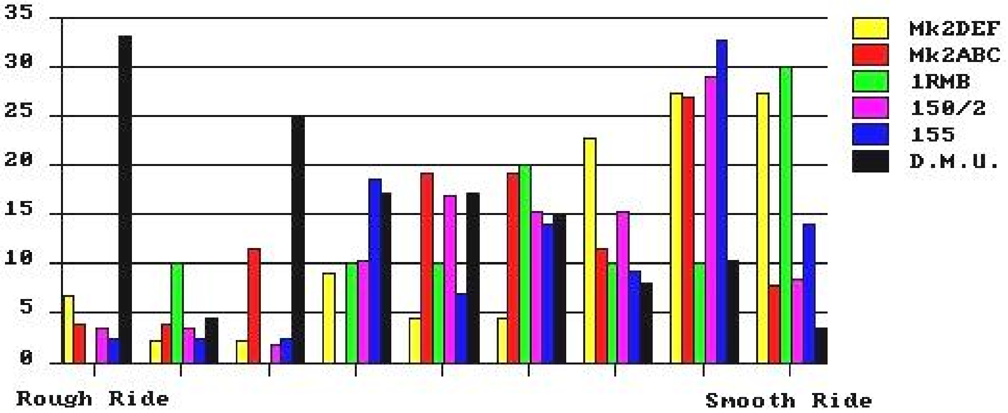


TABLE 7.14: MEAN RIDE SCORES FOR EACH TRAIN.

STOCK	MEAN SCORE
Mark 2 def	6.20
Mark 1 RMB	5.82

150/2	5.70
155	5.56
Mark 2 abc	5.19
Old D.M.U.	4.06

The non-normal nature of the data and the considerable differences in variance between the groups (see table 7.10) mean that the analysis was done using a (non-parametric) Kruskall-Wallis ANOVA. The number of paired comparisons to be done, meant that a series of Mann-Whitney tests would lead to an unacceptably high type I error. A null hypothesis was generated, "There is no significant difference between the scores of the six groups".

Kruskall-Wallis Statistic (corrected for ties) = 42.422

Significant @ 99.5% with 5 d.f.

The null hypothesis was easily rejected, meaning that it was clear that some of the paired comparisons were significant. All possible paired comparisons were now done using the non-parametric equivalent of Scheffe's technique (Marasculio and McSweeney (1977)). The table below shows the results.

TABLE 7.15: COMPARISONS OF RIDE PERCEPTIONS BETWEEN TRAINS.

	2 def	2 abc	150/2	155	1 RMB	Old D.M.U.
2 def	-		-	-	-	-
2 abc	N/S	-	-	-	-	-
150/2	N/S	N/S	-	-	-	-
155	N/S	N/S	N/S	-	-	-
1 RMB	N/S	N/S	N/S	N/S	-	-
Old D.M.U.	99%	N/S	99%	95%	N/S	-

N/S = Not Significant @ 80%.

The only comparisons that proved to be significant were between the old D.M.U. and most other stock. Two of the other stock did not produce significant results, these are the Mark 1 RMB and the Mark 2 abc.

The Mark 1 RMB finding is probably a result of the small sample size (n = 11) as the result is almost significant at 80% and its mean perceived ride score is the second best. The Mark 2 abc result is more puzzling, as its ride should be same as the other replacement stock (which does show a significant difference with the old D.M.U). This lack of significance infers that the ride on the Mark 2 abc was perceived to be worse than all other replacement stock. However the comparison between the Mark 2 abc and other replacement stock was not statistically significant.

The slight difference between the scores of the Mark 2 abc and other replacement stock, could be due to interference effects. It has already been stated that the Mark 2 abc stock was in poor repair and dirty. The Mark 2 abc stock was also noisier than the the Mark 2 def stock, though it should not have been noisier than the Sprinters and the Mark 1 RMB. From this it could be suggested, that of the difference in scores between the Mark 2 abc and Mark 2 def

(1.01) approximately (0.5) was due to noise and (0.5) due to the general condition of the stock.

The most likely stock comparison for any future ride valuation exercise would be between the old D.M.U.s and one of the Sprinters. Both these trains have similar noise levels. Looking at the difference in the ratings of these two (approximately 1.50) we can suggest that approximately 33% (0.50/1.50 x 100) is due to interference effects. This compares to a worst case estimate of 42% produced in the earlier work (chapter six).

If a difference in engineering ride between inappropriate stock is compared interference effects could be much greater. These issues emphasise the careful selection of stock for any comparisons. It is also clear that the effects of noise must be considered when comparing ride between stocks.

5.7. Linearity of Ride Perceptions:

As most of the ride perceptions represent very similar levels of engineering ride, it is difficult to make any conclusions about the linearity of ride perceptions. This problem is made more difficult, as there are two sets of engineering ride measures for each train and there is no accepted way of combining these into a single ride measure.

To establish linearity we need at least three sets of engineering ride measures and their associated perceptions. The nearest we can get to this, are the measures for the two sections of track with the D.M.U. and the measure for the 150/2. These are presented below.

TABLE 7.16: LINEARITY OF RIDE PERCEPTIONS.

SECTION/STOCK	LATERAL	VERTICAL	MEAN SCORE
150/2 Castle Cary - Yeovil P.M.	8.0	15.0	5.70
D.M.U. Westbury - Castle Cary	13.7	37.8	4.36
D.M.U. Castle Cary - Yeovil P.M.	23.0	40.0	4.06

If this data is graphed neither of the two ride measures when plotted against the mean scores is linear. Further both the plots are curved in different directions, suggesting that some combination of the two ride measures may give a linear result. However with this small amount of data, the question of linearity will have to remain open.

6. SUMMARY OF FINDINGS:

Data was collected on passengers' perceptions of ride for six types of rolling stock, representing two levels of engineering ride.

Two surveys were done, the first involved the collection of information on an old D.M.U. service. Data was collected on personal characteristics, the nature of the trip and the travel experience of the respondent. Perceptions of ride were measured for two sections of track and for the overall journey. Two different techniques were used to measure ride perceptions. An attempt was made to see if the importance of ride changed with the environment.

The second survey involved the collection of ride perceptions for five types of replacement stock, using the most successful techniques from the first survey. Engineering ride measurements were made that represented all types of stock investigated over the relevant sections of track.

The two sections of track considered were of very different type: one was largely continuously welded flat bottomed rail and the other was a mixture of jointed bullhead and jointed flat bottomed rail. The former characterises mainline routes, while the latter is more common on rural routes. A significant difference was expected in engineering ride between the two sections. Measurements showed the expected difference, but it was not as large as predicted (sections 2.2.1 and 3).

The difference in engineering ride between the old D.M.U. and the Sprinters (the most realistic replacements) was of a greater magnitude than that between track sections. The relative size of these differences was reflected in the perceptions of ride. The difference between track sections was not always perceived (one test not significant, other test 91% significant) while that between the Sprinters and the old D.M.U. were (both tests 99% significant).

Two ways of measuring ride perceptions were used. With the During approach passengers assessed ride while it actually happened. With the After approach passengers were asked to assess ride at the end of the section/s concerned (section 4.2). As expected from previous research, the During assessments gave significantly more severe results (a result of hyper-sensitivity).

It was thought that the During assessments would be more reliable than the After ones, but this was only the case where two sections were held in memory (almost significant @ 90%). The After assessment of the last section experienced proved just as reliable, as either of the During assessments. But, the earliest section experienced with an After assessment (involving the greatest recall) proved to be less reliable (almost significant @ 90%) than both During and the most recent After assessment. After assessments are considered the closest to replicating an actual travel decision, the After assessments also suffered fewer implementation problems (sections 2.3. and 2.4). The After approach will therefore be used in the future, making assessments of multiple track sections less favourable.

Ride scores were also calculated and compared for individuals' overall journeys, but no significant results were found here (section 4.8). The differences in engineering ride were small between overall journeys, as good and bad sections evened out overall measures.

Mean ride scores were calculated for each train (table 7.14). As expected the old D.M.U. was perceived to be worst. Most stock reported a significant difference in perceived ride from the old D.M.U, though no significant differences emerged between the replacement stock (table 7.15). The mean ride ratings were spread in a way that suggested, a small interference effect from noise and overall condition of stock. Care must therefore be taken to control for differences when comparing engineering ride between stock.

The reliability of the ride assessments was found to have significantly improved in the second survey, this is thought to be the result of improvements in the questionnaire (table 7.10). There were some differences in the age and sex structure of each stock's sample (table 7.13) this may make the assessments of the second survey slightly more lenient.

Respondents were found to have some experience of other lines in the area, but this is probably not sufficient to allow comparisons of engineering ride on different routes in a later stage (table 7.7).

It was found that nearly all passengers made their decision to travel while at home. Previous research has suggested that the environment in which a decision is made can affect the outcome of the decision. To see if ride was affected in this way, respondents were asked about the importance of ride both on the train and at home. No significant environmental effect was found. This may be partly due to imperfections in the method, but even so this partly removes one of the justifications for a household valuation survey (section 4.5).

It was found that respondents are consistent in their ride assessments. Someone who assesses one section of track more critically than the mean, is likely to assess any other section of track more critically than the mean and vice-versa (section 4.4). There is an indication that respondents sat at the end of an old D.M.U are more critical of ride than those sat in the centre, this corresponds with engineering ride measures.

Unfortunately no firm conclusions can be drawn about the linearity of ride assessments (section 4.10).

It was thought that the duration of a respondent's trip may affect their perceptions of ride, making them more or less critical. But no significant results emerged (section 4.6).

A number of personal and other characteristics were found to be significantly related to ride perceptions (section 4.1). These broadly correspond with those found in an earlier investigation (chapter six). Experience of other lines was found to make passengers slightly more critical (86% significant). People aged 16-24 seemed more critical than other age groups, generally older respondents were less critical (83% significant). Passengers with more experience of standing were more critical of ride (98% significant). A respondent's perception of ride is not surprisingly closely related to their assessment of ride's importance.

It was thought that non/infrequent users may be differently critical of ride compared to other passengers. This would mean that special attention would have to be paid to the former group in generating values of ride. But the relationship between the frequency of rail travel and ride assessments did not prove significant.

Originally it was intended to produce engineering ride measures for all types of stock, on all sections of the route. Thus in a valuation exercise based on direct experience of ride, each respondent could be asked where they had travelled so that an engineering measure could be directly associated with their valuation. But difficulties with measurement (section 5.4) meant this was not possible.

All ride valuations will therefore have to be based on the engineering measures taken so far. A stock change valuation would have to be related to the levels of engineering ride measured between Castle Cary and Yeovil Pen Mill. As this section is of relatively poor quality it could slightly exaggerate the difference in engineering ride associated with a valuation (shown in chapter two).

The limited number of engineering measures could cause more serious problems if the differences in engineering ride between track sections were to be valued. However there are now good reasons for abandoning this approach. These reasons are discussed in the next section.

Further the main aim of this thesis is to produce values for a change in ride that can be related to engineering ride, rather than concentrating on engineering measures themselves. If the ride values produced by this thesis were to be applied, more engineering measures could easily be taken by British Rail.

6.1. The Problems of the Track Sections Approach:

6.1.1. Size Of Changes In Engineering Ride:

It is clear that the difference in engineering ride between the two track sections examined was not always detected by respondents: while that between the old D.M.U. and the 150/2 Sprinter was. It is useful to look at the differences in engineering ride for both the track sections and the above stock comparisons. The two track sections tested were Westbury to Castle Cary and Castle Cary to Yeovil Pen Mill.

Between Tracks: Lateral = 9.3 Vertical = 2.2.

Between Trains: Lateral = 15 Vertical = 25.

The engineering ride difference between the stock was considerably greater than the difference between the track sections. However as mentioned earlier (sections 2.2. and 3) the smoother track section contained some poor track where loops to intermediate stations were encountered. With hindsight a smoother section of route could have been found, to compare with the roughest section of the line between Castle Cary and Yeovil Pen Mill. As an example, the difference between the best (Trowbridge to Westbury) and worst track section on the line was:

Between Tracks: Lateral = 15 Vertical = 7.5

This is clearly an improvement over the sections originally used, being much closer to the difference between the two types of rolling stock. But the vertical difference is still only a third of the difference between the Sprinters and the old D.M.U.s. Ideally a valuation exercise would be based on two adjacent track sections - maximising the number of passengers who have travelled over both sections and reducing recall interference. The biggest difference between two adjacent track sections (Bruton to Castle Cary and Castle Cary to Yeovil Pen Mill) was:

Between Tracks: Lateral = 12.5 Vertical = 8

Although it would have been possible to find two adjacent track sections with a greater difference in engineering ride; the difference may still not have been great enough to be discernible to respondents. This can be shown by a rough calculation. The result is only approximate as there is no correct way of combining lateral and vertical ride measurements.

The mean perceived ride scores for the 150/2 Sprinter and the old D.M.U. were 5.82 and 4.06 respectively. The mean perceived ride scores for the two track sections, with the old D.M.U. were 4.36 and 4.06. The difference in engineering ride between the most different adjacent track sections was approximately 2.3 times that between the sections used in the research. So we would expect the difference in perceived ride scores to be approximately 2.3 times greater than with the original track sections, that is 0.70 (0.3 x 2.3).

Adding this difference to the mean score for the roughest section would give a prediction of the score for the smoother section (4.06 + 0.70 = 4.76). This higher score is still below that of the worst improved rolling stock (Mark 2 abc) and so the difference between the extreme track sections, may still not always be perceived by respondents. It may be that the only way to find a large enough difference in track quality is to use specially altered track.

6.1.2. Implementation Of A Track Section Approach:

A comparison of engineering ride between two track sections does not obviously suffer from interference effects. So if a significant number of passengers regularly travel over both sections being considered and there are good memory cues at the section ends, a ride valuation approach could be attempted.

But there is a further problem with the track section approach - the difficulty of translating it into a ride valuation experiment. The only effective way of doing this would be to get respondents to compare hypothetical trips, where the ride was all like section X or section Y. Such trips would not be close to the respondent's previous experience and this is likely to cause significant inaccuracy in the results.

It has been shown that as hypothetical situations deviate from the respondent's experience, the accuracy of the response is reduced (for example, Kroes and Sheldon (January 1988), Fowkes and Wardman (January 1988), Barnard et al (January 1988), Bradley (January 1988)). The difference in engineering ride between stocks is much closer to the experience of the respondent.

On the basis of this discussion, valuation approaches based on the differences in engineering ride between track sections are not to be developed further. By eliminating the track section approach, the remaining approaches will all be valuing the same change in engineering ride making the results more comparable.

7. IMPLICATIONS OF RESULTS:

Before this stage of interviews was completed a number of possible ways of establishing a ride value had been developed (chapter three). The practical value of each of these approaches was dependent on a number of unknown factors. This and the previous chapter have produced firm indications on a number of issues that will now allow a series of valuation exercises to be developed. It is now clear that the number of valuation approaches can be reduced significantly. The extra work involved with household surveys cannot be justified. It is also likely that any approaches involving the difference in ride, between track sections are no longer feasible. This leaves the following approaches, which are developed in the following chapter.

a). Valuation of the mean positions on the textual ride scale using a trade-off analysis.

b). Valuation of ride after a significant change in rolling stock (that is old D.M.U. by Sprinter) seeing how many passengers would be lost after a reduction in engineering ride to the previous level.

c). Valuation of ride with different stock running together using a trade-off analysis.

Chapter Eight Valuation of Ride - Method

1. INTRODUCTION:

From the previous results (chapter seven) we were left with three ways to determine a value of ride. Two of these approaches involve a trade-off analysis and so share many characteristics. The first trade-off illustrated the ride levels, by referring to the ride of the current train and the ride of the (previous) train shown in a photograph (the photograph trade-off). The second trade-off illustrated the same ride levels, using positions on a textual ride scale (textual scale trade-off) developed earlier in the research (chapter five). The third approach was more straightforward and simply asked respondents whether they would still have travelled today, if ride was reduced to its previous level (stock replacement approach).

It had become clear in chapter seven that all ride valuations would be done on-train. The implementation of each approach is considered in detail below. The pre-testing and piloting of each technique is also considered and finally the results of each valuation exercise are presented.

2. BASIC SURVEY DESIGN:

2.1. Pre-testing and Piloting:

A series of pre-tests were done on members of staff and students at Cranfield to ensure that piloting would commence with no obvious flaws. All three valuation approaches were tested in this way. Pre-test respondents completed the tasks by imagining they were on a Thameslink service to London (which they would be familiar with) as this represented a travel environment close to a Sprinter. All three approaches were then piloted on the Bristol-Weymouth line on September 29/30th 1989 (Friday and Saturday).

Many aspects of the research design were changed as a result of this testing - these are described in the following sections.

2.2. Selection of the Study Environment:

Both the pilot and the valuation interviews were carried out on the Bristol to Weymouth line, as this was one of the few Provincial lines that involved long journeys and where Sprinters had recently replaced old D.M.U.s. By conducting the valuation exercises on this line the previously taken engineering measures (chapter seven) would be applicable to the changes in engineering ride being valued. The author also had considerable experience of this line and it was convenient for the surveys.

Class 150/2 and 155 Sprinters were introduced on the selected line from May 1989, the valuation interviews took place in October 1989. All approaches therefore measured the value of a reduction in ride, to the level of an old D.M.U. (chapter three).

Valuation interviews were executed on both types of Sprinter, as the difference in ride quality between them was not found to be significant (chapter seven). Although there was some difference between the Sprinters in other attributes, most of the main ride contaminants (for example, noise) were very similar between trains. Sampling would have become even more difficult and time consuming, if only one type of Sprinter had been concentrated on.

The response rate for all three approaches was very high. The fact that the survey was being done by a student for a Ph.D, seemed to encourage people to participate and perhaps give more honest answers than if it was a purely commercial survey. A number of comments were made to this effect.

2.3. Selection of Services:

Passengers were sampled on all practical Sprinter services during weekdays. Weekends were avoided as many people were changing trains (making them ineligible for the study, section 2.4.1). All trains on the weekday winter timetable were sampled at least once. An effort was made to sample evenly throughout the day and to avoid repeatedly interviewing on the same services. This was done, both to ensure a representative sample and to avoid approaching the same passengers.

The only exceptions to this were the last two trains out of Weymouth, as it was not possible to return if these trains were sampled.

As with previous surveys trains were also avoided where passengers' journeys would not have been as expected. People were therefore not interviewed: on trains running more than 10 minutes late, or where facilities were out of order (catering or toilets). By following this procedure: the response rate was kept high, the exercise was likely to be more accurate and taken more seriously.

To conduct the trade-offs effectively, the interviewer needed somewhere to sit, without invading the personal space of the respondent. The cramped unidirectional seating layout of the Sprinters made this difficult. It was therefore not possible to interview on very busy trains (generally in the peak, north of Trowbridge).

2.4. Selection of Passengers:

Originally it had been intended to select passengers using random numbers. Each passenger on the train was to be allocated a random number, only those passengers with numbers above some cut-off (depending on the sampling rate) would have been selected for interview. However so few passengers proved eligible for the survey that this form of sampling was abandoned.

2.4.1. Passengers Changing Trains:

Passengers were only interviewed if they did not have to change trains, "Today". This was done as passengers would be remaking their travel decision, during the valuation exercise, on the basis of a change in today's ride quality. In the pilot (done on a Friday evening and a Saturday) it proved extremely difficult to find passengers who were not changing trains; approximately 70% of the passengers approached in the pilot were ineligible for this reason. A number of ways were considered to reduce this loss of potential respondents.

It would have been possible to redesign the approaches, so that people who were changing trains could have been included in the sample. But to cater for passengers who had experienced several different trains (levels of engineering ride) that day, would have made the valuation exercises very complex or too unrealistic.

Changing passengers could have been asked to consider only the current part of their trip. But the results would have been much more difficult to interpret and their assessment of the current ride would have been contaminated by the ride on other lines they had experienced that day.

The approaches could have been altered so that they considered totally hypothetical trips - though previous research has shown such designs to be significantly less accurate (section 5.3).

Finally a new line could have been selected, where there were fewer people changing. Such a line would need to have been associated with similar engineering ride levels as the Bristol to Weymouth line and long enough journey times to allow trade-offs to be completed. As mentioned earlier (section 2.2) there were few such lines available.

It was therefore considered unwise to redesign the approaches and impossible to change the line. The solution adopted was to do the final survey on weekdays, when there was expected to be fewer changing passengers. The real survey was also done during the winter timetable (with fewer services) further reducing the numbers of interchanging passengers. These actions meant that the proportion of passengers lost because of this filter, fell to only 30-40% in the real interviews.

During the pilot, it seemed that respondents were finding it disconcerting to be given the introduction to the survey and then left because of the filter question. As a result of this difficulty, rejected respondents were asked some further questions (and the answers written down) to make the process seem less sudden. Respondents were usually asked where they had changed and where they got on the train originally. Obviously using this approach increased the time taken to sample respondents.

2.4.2. Approaches Referring To The Previous Train:

All potential respondents were asked whether they could remember travelling on an old D.M.U. (shown in a photograph). If people could not remember the old train, the interview was terminated, as respondents had to recall the previous train to be able to value its engineering ride. During the interviews, this filter question was expanded, to also ask whether the previous ride was any rougher or smoother than today; as some people could remember the old train, but not its engineering ride. Some passengers even suggested, that the old D.M.U.s were smoother than the Sprinters. All respondents who could remember the previous level of engineering ride were asked to value it (chapter four).

Newness effects have been contained, or even eliminated, as the replacement Sprinters were not introduced from new - they were transferred from other lines. There was also a six month gap between the first introduction of the cascaded trains and the valuation interviews.

It is believed that other interference effects were controlled by comparing ride between similar types of stock (chapter seven). It would not have been wise to compare perceptions of ride between an InterCity 125 and an old D.M.U. Ideally only Class 150/2 Sprinters would have been selected for the interviews, as these present the worst Sprinter passenger environment (in some ways poorer than the old D.M.U.s). Such a comparison would have helped to further reduce interference effects. However it did not prove possible to concentrate solely on one type of Sprinter. This may slightly increase the variation in the results (section 2.2).

It was noted when using the photographs to refer to the old D.M.U.s, that some respondents seemed to have difficulty in concentrating on the ride of the old train. This was less of a problem in the depth interviews, completed earlier in the research (chapter four). The most likely explanation of this difference would be recall difficulties. The depth interviews were done quite soon after the introduction of the new stock; with the valuation interviews, the new stock has been gradually introduced over the last year. It is also likely that users of the London to Bournemouth line travel more frequently, increasing their experience of the old stock.

The photograph based trade-off was intended to be applied, where Sprinters and old D.M.U.s were running side-by-side in sufficient numbers for everyone to have recent experience of both. But this would have meant conducting these interviews on a different line from the other approaches - greatly increasing the time taken and making the results incompatible. This means that some degree of recall error has been introduced into this trade-off's results, though there will now be smaller Newness effects and fewer teething problems (associated with new services).

With both, "Previous train" based approaches the old D.M.U. was always referred to as, "The train in the photograph", to avoid using emotive words, like old and new. Each respondent was shown photographs of inside and outside of the old trains to assist recall. The photographs were carefully taken during previous surveys, ensuring that the units were portrayed in a neutral condition. But even with the care taken over these pictures interference effects may be accentuated by the images of attributes other than ride. It is believed that any interference effects generated by these photographs is likely to be smaller than the errors caused by leading descriptions and poor/incorrect recall.

2.4.3. Other Issues:

Passengers were not interviewed if they were travelling on a free pass; as the idea of paying more or less for the current trip would be unrealistic. Care had to be taken in the phrasing the free-pass question, so as not to infer that the passengers was travelling illegally. Very few people were found to be were travelling on passes.

To make sampling more effective, all three types of questionnaire were carried and those with the most restrictive sampling criteria (the previous train based techniques) were given out first. So if a passenger could remember engineering ride on the old D.M.U, he was given either a stock replacement questionnaire, or a photograph based trade-off. If a passengers could not remember the ride of the old D.M.U, they were given a textual scale trade-off.

2.5. Sampling Bias:

2.5.1. The Need To Identify Bias:

When conducting research into consumer preferences, we would ideally like a sample that is representative of all potential product users (the whole population). Time and money constraints mean that this is not usually possible. It is important, as far as possible, to identify any groups that are under-represented in the sample, to see how this may affect the results of the study. There are a number of such groups in the valuation sample.

2.5.2. Sources Of Bias:

Firstly, it was decided in previous stages (chapter seven) that a household based valuation of ride would not be attempted. This means that infrequent rail users (mostly car owners/users) will be under-represented. Previous work (chapters six and seven) suggests only a weak positive relationship between the severity of ride ratings and frequency of rail use. Attempts to produce a model based on personal and trip characteristics (chapter nine) showed no significant relationship between respondent's ride valuations and their frequency of rail use.

Passengers making journeys that involved changes were not included in the survey. It may be that these passengers were more fatigued and thus had different values of ride. No direct evidence has been produced in the research, to confirm or deny this. What can be said is that the relationship between trip length and ride assessments (chapter seven) as well as that between trip length and ride value (chapter nine), proved insignificant.

Passengers who could not remember the ride on the old D.M.U.s were excluded from the previous train based approaches. To make sampling more productive (section 2.4.3) these passengers were often given textual scale trade-offs. These people may have been less frequent travellers with different views. Previous work (chapters six and seven) has suggested that infrequent users may be slightly less critical of ride, though no significant relationship was found between people's ride values and their frequency of rail use (chapter nine). The textual scale trade-off sample probably has a greater representation of infrequent users than the two previous train based approaches. The results of the textual scale trade-off may therefore be slightly less critical than the others.

Passengers travelling on free passes were not interviewed, but this is unlikely to affect the results as very few were encountered.

Those who refused to be interviewed may have had different opinions. Intuitively one would expect such people to be more critical of ride, but no evidence is presented for this. As so few people refused to be interviewed this effect can be safely ignored.

The last two trains out of Weymouth were not sampled. But very few people used these trains and so it is not expected to have affected the results.

Weekend travellers were not sampled. One would expect more leisure trips during weekends and also longer trips characterised by changes of trains. As stated previously no evidence can be found for the effects of interchange on passengers' ride values. The relationships between trip length and ride assessments (chapter seven) as well as that between trip length and ride values proved insignificant (chapter nine). No significant relationship was found between passengers' ride values and their trip purpose (chapter nine). However there was an indication (chapter six) that passengers on work trips are more critical of ride.

Passengers were not interviewed when the train was crowded. Peak users, who are likely to be slightly more critical of ride (chapter six) would be under-represented because of this. This bias should to some extent counteract the one discussed previously.

The stock replacement questionnaires tended to be given to passengers on busier trains and making shorter trips. This would suggest that the ride assessments generated from this exercise were slightly more critical than the others.

2.5.3. Conclusions:

The difficulties of sampling meant that the planned sample sizes of a hundred each for the trade-offs and two-hundred for the stock replacement approach had to be halved. The original sample sizes were clearly impractical under the circumstances.

From the discussion above it is clear that there are a number of biases in the valuation samples. However from previous work (chapters six and seven) few of the biases are likely to have significantly affected the results. Although there are indications that certain groups are more/less critical of ride, these biases seem to counteract each other and none of the groups showed significantly different values of ride (chapter nine).

2.6. General Questionnaire Design:

Great care was taken with the design of all three questionnaires to ensure that they looked interesting and professional. All questionnaires were set out on headed paper and printed onto coloured card. Each approaches' questionnaire was coloured differently, so that they could be easily distinguished by the interviewer. Printing onto card also made it easier for respondents to write without support, ensuring that answers could be easily interpreted during coding.

Once selected all respondents were asked preliminary questions about their: age, sex, trip purpose and frequency of rail use, this information was used later in the construction of a model.

3. ECONOMIC BACKGROUND:

3.1. The Utility Model:

The additive utility model illustrated below (based on those used conventionally, see chapter three) was adopted as the foundation for all three valuation exercises. It is hypothesised that an individual gains some benefit from moving to another location, from this is deducted the disbenefit associated with travel to the other location. For example, someone may gain X utils from going to the cinema in the next town, but loses Y utils in travelling there. If the value of X - Y exceeds the value of the alternative courses of action (for example, watching T.V. at home) that trip will be made. It is therefore assumed, from basic microeconomic theory, that the consumer adopts the solution which maximises his utility (chapter three).

The disutility associated with the travel involved in any course of action, is an additive combination of the utilities associated with the attributes that make it up.

FIGURE 8.1: THE UTILITY MODEL.

$$\text{UTILITY} = f \left| \begin{array}{l} \text{- Ride, - Fare, - Journey Time,} \\ \text{- Other Attributes,} \\ \text{+ Benefit from moving} \end{array} \right|$$

For the analysis to be successful we must be able to control certain factors in the model. "Other attributes" are held constant by stating that all unmentioned factors are the same as the current trip. "Benefit from moving" is controlled across the alternatives, by only comparing different ways of making the same trip. Thus in the trade-offs by examining the way overall utility varies along with the levels of ride, fare and journey time, we were able to determine the relative importance (value) of each attribute in the choice process.

3.2. The Relationship between Demand and the Values of Attributes:

Two forms of ride valuation approach have been developed, each of these produces different types of result. The stock replacement approach gives a result in terms of the change in demand for a service that would result from a change in ride quality. The trade-off approaches generate results in terms of the financial value of individual attributes.

At this stage it is therefore important to clarify the relationship between the demand for a good or service and the value of changes in individual attributes that make up any good or service.

From basic economic utility theory we can state that demand is the quantity of a good/service (choice) that will be bought at any given price. Further that a consumer will demand the choice which yields the maximum amount of utility. The utility associated with any choice is some combination of the utilities associated with the levels of the various attributes that make it up. Thus any change in the level of an attribute is likely to change the overall level of utility associated with that choice. The size of this effect depends on the value consumers place on changes in that particular attribute. Thus if an attribute is changed that has a high value with consumers, the overall level of utility and thus demand for that particular choice will be affected more than if an attribute with a low value were changed.

4. VALUATION OF RIDE AFTER A SIGNIFICANT CHANGE IN ROLLING STOCK:

This method was explained in detail earlier (chapters three and five). This approach attempted to establish a value of a change in engineering ride, by seeing how many of the current passengers would be lost if ride quality was reduced to some specified level. Respondents were asked to think back to when they made the decision to travel today. They were then asked whether they would still make today's trip, if they had expected the ride to be like that of an old D.M.U. This question was phrased in a way that should reduce interference effects (chapter five). The proportion that would no longer make today's trip, would allow us to produce an indication of the value of ride.

To make the exercise as valid as possible there are a number of factors that had to be accounted for.

For this exercise to be successful we needed a line where Sprinters had completely replaced a fairly standard fleet of old D.M.U.s. It was therefore clear to respondents' what the old trains were. Ideally the replacement would have been completed two or three months before the interviews, reducing Newness effects and ensuring that all respondents expected to travel on a Sprinter, "Today". If the replacement happened too long ago recall error would cause the assessments to be unreliable. There is clearly a trade-off here between recall error and Newness effects.

The line selected for the valuation broadly fulfilled these needs, though there were one or two old D.M.U.s still running occasionally. As these old trains only ran to cover for failures and on small sections of the route very early in the morning, this is not expected to have affected the results.

The six month gap between the introduction of Sprinters and the valuation interviews was probably too great, as some passengers had difficulty in recalling ride on the old trains. This problem was accentuated as most respondents used the service infrequently. This infrequent use may mean that not all respondents had expected to travel by Sprinter. Nevertheless the long lag after Sprinter introduction should mean that most people had expected to be using the new trains. Anyone who did complete the questionnaire and had expected to be travelling on the old D.M.U. would probably have stated that the reduction in ride quality would not affect their trip - which is a valid response.

The stock replacement approach was so straightforward that the questionnaires were designed for self-completion. Each respondent was left a photograph of the old train and a questionnaire, these were collected later in the journey. After rewording the main question during the pre-tests this approach was successful.

The low expected value of ride (chapter two) meant that only a small proportion of the sample would no longer make today's trip. This meant that a large sample was required for this approach to be successful. As we were dealing with a large change in engineering ride it was decided that a sample of between a hundred and a hundred and fifty would produce an accurate result. The questionnaire was very straightforward and suitable for self-completion (appendix eleven) so a sample of this size was considered attainable.

5. PRINCIPLES OF TRADE-OFF ANALYSIS:

5.1. Foundation:

Trade-off analysis is based on the premise that people gain utility from consuming goods and services, in this case travelling by rail. This utility associated with a good/service is the sum of the utilities associated with its various attributes, for example: journey time, comfort etc. Individuals' resources are finite and they thus have to choose between alternative goods and services. Each person chooses by selecting the good/service that maximises their utility (Laidler (1981)). From this it follows, that it should be possible to simulate the choice process under controlled conditions and extract from individuals the utility value associated with any attribute of a good/service.

5.2 Implementation:

The choice process is usually simulated, by giving selected individuals a series of hypothetical goods/services (options) each with different levels of attributes. Individuals are then asked to choose between the options, causing them to trade-off attributes for each other. In this thesis the options are a series of rail journeys and ride quality is one of the attributes to be traded by respondents. As this form of stated preference design is more efficient than revealed preference, samples can be smaller than would otherwise be the case (Bradley (January 1988)).

Respondents can compare options by: choosing between pairs (choice), giving a rating to each (rating) or ranking the options in terms of favourableness (ranking). If rating or ranking methods are used, we have to assume that the highest rating/ranking constitutes first choice (Louviere (January 1988)). The latter two methods do not produce actual choices and so it is argued that the results may be less valid than sets of choices made by respondents (Louviere (January 1988)). It is believed that ranking and choice are the easiest for respondents, making them more accurate (Green and Srinivasan (September 1978)). But choice tasks are more difficult to design and less efficient than rating or ranking. Rating produces more information than ranking and when coupled with a continuous (utility) scale, the results are more amenable to regression. Although regression can be applied to ranked data (with rank being the dependent variable) the results are not as valid as with a (metric) rating dependent variable (Aaker and Day (1986)).

5.3. Design:

To design a trade-off, one initially selects the attributes and levels of attributes. The next stage is to decide on the experimental design - the number of different options (combination of attributes) to be assessed. The design can either use repeated measures (with each respondent assessing a number of options) or cross-sectional measures (with each respondent assessing only one option). A substantial amount of among-person variation is expected in consumer preferences (Green and Srinivasan (September 1978)) so trade-off analysis is generally carried out at the individual level. With repeated assessments a model is calibrated for each individual, the form of the model is generally the same for all individuals though the parameters vary between them (Green and Srinivasan (September 1978)). The resulting values are aggregated to represent the sample (and estimate the population). With a cross-sectional approach a model is calibrated across the respondents.

Green and Srinivasan (September 1978) describe a number of possible experimental designs and the difficulties associated with them. Obviously the more data that is collected, the greater (cet.par) the accuracy of the results. For the best results we must maximise the n/T ratio, where n is the number of options evaluated and T the number of of attributes (Green and Srinivasan (September 1978)).

Ideally a full factorial (all possible combinations are presented) orthogonal (no correlations between attributes) design is chosen, as this maximises the accuracy of parameter estimation. But such a design is rarely possible as it requires too much effort from respondents - a design with three attributes each with three levels would require the assessment of (3 x 3 x 3) twenty-seven options. The number of comparisons required from each respondent, must be kept to a minimum to avoid boredom/fatigue (resulting in lower accuracy) and to keep the response rate high (Green and Srinivasan (September 1978)). For these reasons most trade-off research uses fractional factorial approaches, which with careful design can still be orthogonal.

As far as possible the design should present options that are not dominated. If the result of a comparison is dominant (with all attributes the same or better, or all the same/worse) little is learnt from the exercise. With a rating approach dominant options compress other ratings on the scale, reducing the difference relative to the measurement error (Bradley (January 1988)). It has been found that choices between competitive options are more interesting for the respondent, resulting in more accurate responses (Bradley (January 1988)).

To achieve an orthogonal design, it may be necessary to add some dominant combinations into the exercise. This also allows a test for irrational responses (Barnard et al (January 1988)). But care must be taken with this, as orthogonal designs can present unbelievable combinations of attributes - causing respondents to make inaccurate assessments.

To obtain valid responses, the attribute levels must be close to the respondent's experience, or be related to the current trip (Kroes and Sheldon (January 1988), Fowkes and Wardman (January 1988), Barnard et al (January 1988), Bradley (January 1988), Green and Srinivasan (September 1978)). It is also important that the options are realistic (Bradley (January 1988), Green and Srinivasan (September 1978)) or the results will not be valid. Respondents would, for example, find it difficult to believe that an InterCity 125 has a greater journey time than an old D.M.U.

Contrasts between options should be typical, so that results can be applied to other populations. Options should be presented to respondents in random order to counteract order effects (Barnard et al (January 1988)).

Ideally face-to-face interviews should be used, as the tasks can be complex and this allows the survey form to be customised (Kroes and Sheldon (January 1988)). It has been found that respondents find options with more than three attributes difficult to understand and the results may be unreliable (Barnard et al (January 1988)).

A design will be more efficient if the attributes are presented at levels at which respondents will be willing to trade (Bradley (January 1988)). It would be no use asking a respondent to trade a 50% increase in fares, for a 5% reduction in journey time. Attribute values should be rounded to the nearest five, as respondents have been found to report cost/time differences in these units. Preliminary work is important to determine attribute levels and choice contexts (Bradley (January 1988)). It is important to minimise complexity, as people give the most reliable results when assessing changes in only two or three factors simultaneously. Generally the more complex the exercise, the greater the error (Bradley (January 1988)). Greater complexity tends to to cause respondents to abandon trading choice rules (Bradley

(January 1988)). Bradley suggests starting respondents with the simplest tasks as a warm up (perhaps not using the results) but this is likely to lead to order effects or fatigue/boredom.

With repeated measure designs, respondents are usually presented options on a series of stimulus cards - allowing the presentation sequence to be varied with respondents, to counteract order effects. It has been found that the importance of an attribute can be related to its position on the stimulus card (Green and Srinivasan (September 1978)) this order must therefore also be varied. Pictorial based approaches have been found to produce quicker results, as they reduce information overload, make the task more interesting and increase the homogeneity of perceptions across respondents (Green and Srinivasan (September 1978)). However the attributes studied in this thesis are not amenable to pictorial representation. Pictures may also convey unexpected signals to respondents, which may interfere with the attributes being measured (Green and Srinivasan (September 1978)).

Once a design has been finalised it is very important to pre-test it to ensure that it is working as expected (Green and Srinivasan (September 1978)). De-briefing after piloting is important to identify any difficulties with the design (Bradley (January 1988)).

5.4. Analysis:

Utility values are usually extracted using: MONANOVA, Maximum likelihood, or (metric or non-metric) regression techniques - depending on the nature of the data and the availability of software (Kroes and Sheldon (January 1988)). Metric regression is the most readily available and can be used, if a utility measure is generated as a dependent variable. Models can be estimated for individuals, as well as groups of individuals, if respondents do enough trade-offs.

The use of repeated measures can cause statistical problems (Kroes and Sheldon (January 1988)). For example, using regression with repeated measures (Groebner and Shannon (1985), Aaker and Day (1986), Open University (1976)) can result in hetroscedasticity (where the variance of assessments is not constant) and autocorrelation (where residuals are correlated) producing inefficient and/or inaccurate parameter estimates. Making inferences about the population from the standard F and t tests is not appropriate with repeated measures; as repeated assessments from one individual will show less variation than a group of independent assessments from several individuals (on which the tests are based). This incompatibility would make the results seem more accurate than they really are. The difficulty of making population estimates can be avoided when a number of individuals give repeated measures (as is usually the case). When a number of individuals are involved the mean and standard deviation of the individuals' estimated utilities can be calculated - from these population estimates can be obtained. This technique has been found to work very effectively in the simulations (section 6.4).

In choosing between techniques, "As a practical matter most analysts use regression analysis to obtain attribute weight utilities since it provides very similar results (to other techniques) and is much easier and cheaper to use than an iterative procedure" (Aaker and Day (1986)). Aaker and Day (1986) also note that studies comparing the alternative methods of trade-off data analysis typically find that the estimated utilities are, "Largely similar".

Cattin and Wittink (July 1989) have recently completed a survey on the commercial use of trade-off analysis - investigating trends since the techniques were first developed. They report median sample sizes of three-hundred (clearly not possible within the scope of this research) and a median number of judgements per respondent of sixteen (not possible with unpaid respondents on a train). The median number of attributes used in the analysis is eight.

Cattin and Wittink (July 1989) state that there is an increasing tendency to use interactive computer programmes, with respondents, to establish attribute values. The use of computers provides: more flexibility, a more interesting task for the respondent, built in checks for consistency, and immediate results. However the cost of the equipment means computerised data collection would be unsuitable for this research. There has been a move away from ranking towards rating approaches and an increasing reliance placed on regression for the analysis of the results.

5.5 Critique:

Louviere (January 1988) criticises some of the assumptions of trade-off analysis. He makes a distinction between judgement and real choice data, stating that these are not necessarily the same. Difficulties have been reported in mode choice comparisons, respondents may try to justify their current mode, or air grievances by complaining about it (Bradley (January 1988)). But this should not be a problem when comparing alternatives services for one mode. Louviere also states that the utility specification may not be additive - there may be interactions between attributes or non-additivities. The existence of nesting and elimination strategies can invalidate the analysis. Such processes may mean that subsets of choice do not give the same results, as the full set of options. Trade-off models can give false results: as they assume that respondents have full knowledge of the alternatives, there is no bias in the estimated parameters, individuals are transitive and consistent, and there are no other constraints to cause respondents to fail to choose the maximum utility solution.

However Louviere (January 1988) later states that there is considerable evidence of the external validity of these techniques. It has been found that predictions of behaviour from aggregate trade-off models correlate well with the behaviour of aggregates of real people other than those studied.

The possibility of errors mean that it is important to check the accuracy of any designs developed in this thesis. This can be achieved by looking at the results of dominant choices and by comparing any results to those of previous work.

6. THE GENERAL TRADE-OFF DESIGN:

6.1. Introduction:

The previous discussion above and the peculiar circumstances of this investigation, meant that the following procedures were adopted.

6.2. Number and Levels of Attributes:

The only attributes this research is specifically concerned with are ride and fare (as this allows us to establish a financial value of ride). However to establish the validity of the trade-off results, it was decided to include another attribute with a known value. The most effective such attribute, as it is not correlated with ride or fare and has been studied in detail, is journey time. Although an exact value of journey time has not been established a sufficiently well defined range has emerged to allow meaningful comparisons with the thesis results. It was therefore decided to use these three attributes in the trade-off. A further reason for including an extra attribute into the trade-off was so that ride did not stand out too clearly as the attribute we were interested in.

Each option displayed all three attributes, though each attribute did not always vary between options. However it was important that more than one attribute was varied between options (as far as possible) to further disguise our intentions. Further attributes could have been added for this reason, but the increase in complexity would have outweighed any other advantages (section 5.3). Two engineering levels of ride were presented (only two clear levels were established during the previous work - chapter seven) these were based on the 150/2 Sprinter and the old D.M.U.s. There were three levels of fare and journey time - the current levels, one above and one below.

The exact levels of the attributes presented to respondents were determined from the values extracted by previous work in this area. To ensure an efficient design it was important that the levels of the attributes presented caused respondents to trade. It was difficult to do this with ride, as little was known about its value - the only available information was from M.V.A. (chapter two).

We know that the three jumps on M.V.A's ride scale were worth 0.99% (level 1-2), 2.33% (2-3) and 3.69% (3-4) of fare for second class passengers. The major difficulty with these values is that they are not associated with any engineering measures of ride. We therefore had to approximate the value of the change in ride between an old D.M.U. and a Sprinter (on the Bristol to Weymouth route). On M.V.A's scale the ride on an old D.M.U. would be approximately level one, "Rough ride with frequent jerks sufficient to spill drinks from a cup". The Sprinter would be similar in ride to the Mark 3 coaches on which M.V.A's passengers were interviewed and so should fall between levels 2 and 3. This would suggest the ride improvement between the old D.M.U. and the Sprinter is worth roughly 0.99% + 1/2(2.33%) = 2.17% of fare.

The journey time elasticity for Provincial is approximately -0.4, while the fare elasticity is approximately -1 (British Railways Board (1986)). From these elasticities we can say that a given % change in journey time is worth 40% ((-0.4/-1) x 100) of the same % change in fare.

The expected size of the ride value meant that the difference between fare's and journey time's levels had to be very small, for respondents to trade these attributes with ride. Although the trade-off was designed with these values in mind, it was robust to wide variations in these values (shown by simulation, section 6.4).

6.3. Choice of Assessment:

6.3.1. Priority Evaluator:

One way to establish a value for a change in engineering ride would be to use a priority evaluator technique, similar to that developed by M.V.A. (chapter two) but with improved ride descriptions. But M.V.A. applied this technique on an InterCity route, a priority evaluator may be difficult to apply in the context of this thesis. A priority evaluator is perhaps too complex to use on many Provincial lines where journeys are likely to be shorter. Provincial trains are also more cramped and do not generally have tables where which such equipment could be set out.

6.3.2. Chosen Approach:

It was decided to base the trade-off design on an adapted rating approach, which incorporated elements of ranking and choice. Each respondent produced six ratings (section 6.4) based on the current trip and five alternatives. The alternative options were rated relative to the trip currently being undertaken by the respondent (placed at the centre of the scale). Assessments were related to the current journey, as this was relevant to the respondent and fresh in his mind.

The idea of relating all the options to one base case was expected to make the rating task easier for respondents. This design, being nearer a choice, should be more realistic and thus accurate. All five rating scales were closely placed side-by-side allowing respondents to rank the options in order of preference as well as indicating the size of the difference between ranks. The output from this design was also easily analysed with multiple regression. With each respondent doing multiple assessments it was easier to produce effective disaggregate models later on.

Originally it had been intended that each respondent would only compare one alternative to their current trip, so that only a cross-sectional design would be developed. This was thought useful as: it removes the problems of repeated measures in regression, there would be no order effects, less effort would be required of respondents (keeping the response rate high and each assessment more accurate), the exercise could be designed for self-completion and a full factorial orthogonal design could be used.

It was decided to give each respondent multiple assessments allowing both forms of calibration, as repeated measures benefit from less interpersonal variation and need fewer respondents. Repeated measures also allow the development of more effective disaggregate models. Finally, it was found that a semi-factorial orthogonal design was just as accurate as a full factorial orthogonal design (section 6.4).

The above points meant that it was too risky to proceed with a cross-sectional design alone. A cross-sectional calibration was still tried using one different assessment from each respondent. To make the cross-sectional calibration orthogonal, the sample sizes should have been in multiples of five (section 6.4) but in practical terms this was impossible to ensure.

6.3.3. Customisation Of Trade-Offs:

The trade-off options were to be customised for each individual according their current trip and the overall experimental design. This approach was used in the pilot, but for the reasons discussed below it was not possible in the final interviews.

Each respondent's journey time and fare were established from the first part of the questionnaire. Ideally 2.17% would then have been added/subtracted from the fare and (-1/-0.4 x 2.17) 5.43% added/subtracted from the journey time to produce the high/low fare and high/low journey time levels. But the need to rapidly calculate these percentages by hand during each respondent's interview, meant that more easily calculated figures (2.5% fare and 5% journey time) were used in practice. These values were then rounded to the nearest five units (down for fare and up for journey time). If respondents were travelling on period tickets (Rover/Season) the new fares were to be based on the overall value of their ticket, as this would be more relevant to the respondent than the equivalent daily value (this never occurred during piloting).

Using this customised approach it was hoped that the exercise would be more realistic, relevant to the passengers experience and thus more accurate. Unfortunately a number of difficulties emerged during the pilot which meant that customisation had to be abandoned.

The small percentages used and the effects of rounding meant that the customised fare and journey time levels varied little between respondents. It became clear during the pilot that the whole trade-off procedure was too long (the task took each respondent seven to ten minutes). Some respondents were finding it difficult to concentrate over this period. Customising the cards meant that each respondent had to wait while their fare and journey time was read from their questionnaire, trade-off levels calculated and then written onto each of their cards. The need to do this quickly meant that the correct rounding up/down for each attribute was not consistently done. A programmable calculator was sometimes carried in an attempt to make the calculations quicker and more accurate, but it made little difference.

It is important when interviewing on a train to minimise the intrusion on passengers. This is best achieved by working systematically through a coach without having to return for supplies. Enough equipment therefore has to be carried to allow this; the use of customised cards limited the number of trade-off sets that could be carried. Removing customisation greatly reduced the amount of material that had to be carried and the amount of photocopying. Further the clarity of the handwriting (without proper support) could make the trade-off cards more difficult to understand.

For these reasons it was decided to abandon the customised trade-off design. Although it would have been possible to calculate the custom fare and journey time values, whilst the respondent was filling in the questionnaire; this would have meant looking over the respondent's shoulder while they filled in the questionnaire, which is not generally possible without being intrusive.

Standard attribute values were therefore presented to respondents in the final survey. These levels were determined from the average journey lengths and fares established during the pilot (a very small sample). The fare change was plus/minus fifteen pence and the journey time plus/minus five minutes. These levels proved very successful.

Each respondent was still asked their fare and journey time, as this allowed us to establish the percentage change they were considering in the standardised trade-off.

Alternative levels of fare and journey time were stated as differences from the current levels (for example, 15p extra) rather than absolute values (for example, new fare £5). This was expected to make the task easier for respondents.

6.3.4. The Rating Scale:

The above discussion meant that the rating scale was bipolar: the centre of the scale represented a journey as good as today, while the extremes represented journeys better or worse than today. Ratings were produced on a continuous graphic scale similar to those developed earlier in the research (chapter five), but uses three anchors. The central anchor represents the utility associated with the current trip, while the other two suggest an improvement and a worsening.

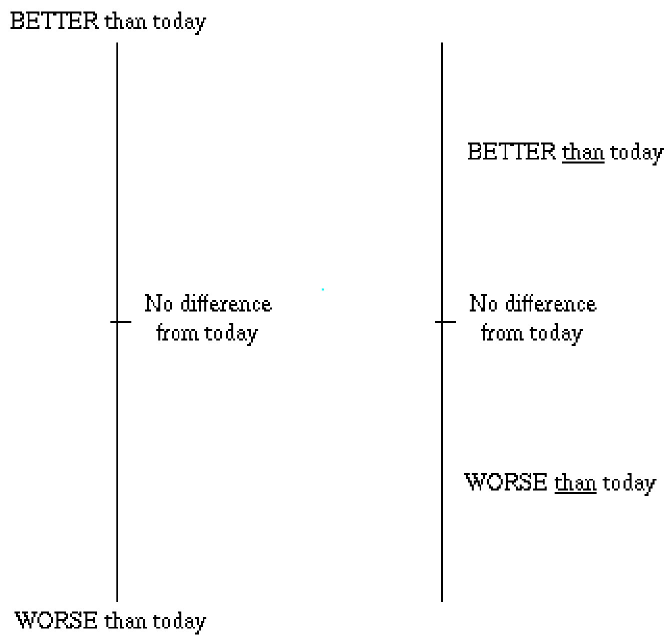
Originally positions on the rating scale were to be coded from minus five to plus five with, "No difference from today" being coded as zero. But it was found during the pre-tests, that respondents were much more precise. During testing interviewees were encouraged to talk about the task while completing it, from this it was clear that some respondents were discriminating by as little as two millimeters. This level of discrimination is equivalent to a forty-five point scale scored as +22/-22.

Pre-test responses were measured at both levels of precision and the results were found to be best when measured at the two millimeter level. Obviously measuring some peoples responses at this level could suggest variation that does not exist. But overall two millimeters appeared to be a sensible level of discrimination and so was used for the final analysis. The fact that such a change could be made so easily after the collection of data is one of the strengths of a graphic scale. It would clearly be impossible to have a discrete scale with enough anchors to work at this level of precision.

The scale's anchors were originally, "No difference from today" in the centre, "Much better than today" at the top and, "Much worse than today" at the bottom. But it was found that the top and bottom anchors were too extreme, causing some bunching of assessments towards the centre of the scale. For this reason the, "Much" statement was removed from the top and bottom anchors.

Experience from the final survey suggests that these anchors would be changed once more, if the technique were used again. The scale that was used is shown on the left in figure 8.2, while the modified version is shown on the right. The scale would have been improved by emphasising, "Than" today as a number of respondents interpreted the scale wrongly and tried to rate today's trip relative to the one on the card. Some respondents took the anchors to mean that better/worse ratings could only be placed at the extremes of the scale. It would therefore have been better to have the anchors on the side of the scale, approximately halfway down.

FIGURE 8.2: ALTERNATIVE GRAPHIC SCALES.



6.4. NUMBER OF OPTIONS AND SIMULATION:

6.4.1. Introduction:

A simulation was developed on a spreadsheet to test the effectiveness of various experimental designs, which are developed in this section. The simulation operated by generating sets of imaginary choice data, that could be investigated using each experimental design. The accuracy with which each design was able to reproduce coefficients (representing the value of the attributes) could then be established.

6.4.2. Experimental Designs:

With three attributes (Ride, Fare and journey time) at two, three and three levels respectively, a full factorial design would involve presenting eighteen (2 x 3 x 3) different options to respondents. The use of attributes with different levels makes this an asymmetric design. The benefits of a full factorial design are that it is orthogonal (with no correlation between the attributes) and that all main and interaction effects can be estimated from the results. But as it was intended to ask each respondent to assess all options, this design was clearly unacceptable. This was especially the case when using unpaid respondents on a train. The full factorial design is shown below:

TABLE 8.1: ORTHOGONAL FULL FACTORIAL DESIGN (A).

RIDE	FARE	JOURNEY TIME
0	0	0
0	0	1
0	0	2
0	1	0
0	1	1
0	1	2
0	2	0
0	2	1
0	2	2
1	0	0
1	0	1
1	0	2
1	1	0
1	1	1
1	1	2
1	2	0
1	2	1
1	2	2

By following the rules outlined by Green (September 1974) it was possible to create a fractional factorial design which was also orthogonal. Using this procedure the eighteen options above can be reduced to six. One of these options (the null option) represents the current attribute levels and all options are rated relative to this. This means that only five options have to be individually presented to respondents, as the null option is automatically rated as zero. Using a fractional factorial design means that interaction effects can no longer be estimated from the results. However in this research we are only interested in finding the size of the main effects and so this loss of information is not important. The reduced design means that some dominant choices are presented to respondents, but these are few and they will allow a test of the validity of respondents' choices. The orthogonal fractional factorial design is presented below:

TABLE 8.2: ORTHOGONAL FRACTIONAL FACTORIAL DESIGN (C).

COLOUR CODE	RIDE	FARE	JOURNEY TIME
R	0	0	0
B	0	1	1
G	0	2	2
Y	1	0	2
NULL OPTION	1	1	1
W	1	2	0

A non-orthogonal fractional factorial design (B) was also developed that avoided dominant comparisons. Design (B) had the highest relative information content, but suffered from having correlated independent variables (B). The three designs were tested by simulating a number of responses on the spreadsheet model.

TABLE 8.3: NON-ORTHOGONAL FRACTIONAL FACTORIAL DESIGN (B).

RIDE	FARE	JOURNEY TIME
0	0	2
0	1	2
0	2	0
0	2	1
0	2	2
1	0	2
1	1	1

6.4.3. The Simulation Procedure:

A number of different simulations were tried, all based on the equation below. This model states that the utility associated with any option, is an additive combination of the utilities associated with the ride, fare and journey time levels that make it up. This form of model is the most commonly used in trade-off analysis and is generally considered to be a valid representation of choice behaviour (chapter three and section 3).

FIGURE 8.3: SIMULATION MODEL.

$$\begin{aligned} Ut = RW \times Ride + RanW \times (Ran - 0.5) \quad & + Fare... \\ & + Journey\ Time... \end{aligned}$$

Ut	= Utility of option.
RW	= Ride value weight.
Ran	= Random number between 0-1.
RanW	= Size of random effect.
Ride	= Ride level -1 or 0.
Fare	= Fare level -1, 0 or 1.
Journey Time	= Journey Time level -1, 0 or 1.

In each simulation a series of options (or assessments) were presented on a spreadsheet. The above equation was used to generate an amount of utility for each level of every attribute within these options. An overall level of utility for each option could then be established. Utility values for each attribute were generated (figure 8.3) based on the level of each attribute presented (Ride), around this level was a variable effect (representing the variation between individuals) produced by random numbers (Ran -0.5). The relative importance of these two effects could be changed by weights (RW and RanW) in the equation.

The utilities generated by the simulation were interpreted in two ways. Firstly data was analysed as if each option was being considered by a different fictitious individual (cross-sectional calibration). Secondly the analysis was done as if one individual had considered a series of repeated assessments (repeated measures). The random effect in the model was of the same magnitude for both cross-sectional and repeated assessments. Although it is likely that a series of repeated assessments would show less variation than a series of cross-sectional assessments; it is difficult to know how much less. Furthermore the main purpose of the simulation was to investigate the difference in the effectiveness of the various designs presented to respondents, rather than to establish the relative merits of a repeated or cross-sectional implementation.

Option utilities were produced for a variety of sample sizes and with different levels of variation. As ride is the least tangible attribute, it was considered realistic to associate its utilities with a greater variation than with fare and journey time. The size of the variation used in most of the simulations, was based on that found during earlier ride assessments. For safety a pessimistic view of the expected variation was taken - the value being slightly greater than the most varied original ride assessments described in chapter seven.

In the final stage of the analysis the total level of utility associated with each option was regressed against the levels of the attributes presented, in an attempt to estimate the utilities associated with each attribute. The effectiveness with which these utilities are estimated indicates the efficiency of the experimental designs.

6.4.4. Simulation Results:

The simulation results were compared on the basis of the R2 values and the standard errors of the estimated coefficients. For successful estimation the former must be maximised and the latter minimised. Tests were done to see the effects of: changes in the deviation of assessments, differences in the relative values of the attributes, the three different experimental designs and the bunching of assessments on the utility scale. The main results were:

TABLE 8.4: SIMULATION RESULTS.

A. ORTHOGONAL FULL FACTORIAL DESIGN:
(108 assessments - cross-sectional)

Mean R2 = 0.66
Standard Error: Ride = 0.172 Fare = 0.106 J.Time = 0.106

B. NON-ORTHOGONAL FRACTIONAL FACTORIAL DESIGN:
(105 assessments - cross-sectional)

Mean R2 = 0.40
Standard Error: Ride = 0.218 Fare = 0.146 J.Time = 0.146

(7 assessments - repeated measures)

Standard Error: Ride = 0.684 Fare = 0.452 J.Time = 0.452

C. ORTHOGONAL FRACTIONAL FACTORIAL DESIGN:
(102 assessments - cross-sectional)

Mean R2 = 0.70
Standard Error: Ride = 0.162 Fare = 0.100 J.Time = 0.100

(6 assessments - repeated measures)

Standard Error: Ride = 0.624 Fare = 0.384 J.Time = 0.384

The non-orthogonal design (B) was clearly the least effective at estimating the attributes' coefficients: this is a result of multicollinearity (correlations between the attributes). It was also apparent that the smaller orthogonal design (C) was just as effective as the full factorial design (A). Only the fractional factorial designs can be realistically used with repeated measures, which makes the orthogonal design (C) appear the most effective.

All the changes made during the simulations produced the expected results. The bunching of assessments (represented by small differences between the utilities associated with each option) was found to significantly reduce the quality of the results and it is therefore important to ensure that anchors on the utility scale cause assessments to be spread throughout the scale's length. All the designs appeared robust and were able to return realistic coefficient values even with ride at a quarter, or four times the value found by M.V.A.

It should be remembered that these were only simulation results involving a limited number of runs (typically three to five). However the small variations in results between runs, suggest that the general pointers derived from the results were sound.

From the above results and the previous discussion, it was decided to adopt the orthogonal fractional factorial design (C) for the trade-offs. This design kept the trade-off exercise manageable for respondents, while the results appeared as accurate as the full factorial design. By using this fractional factorial design, the same data can be calibrated in cross-sectional and repeated ways. Although it would be possible to develop a design with a greater information content, this would either exhaust the respondent or be non-orthogonal and (from the results of the simulation) this is clearly less accurate than an orthogonal design.

6.5. Trade-Off Implementation:

6.5.1. General:

The complexity of a trade-off approach and the desire to obtain more than one rating from each respondent (while avoiding order effects), meant that data had to be collected in the form of an interview.

Respondents were initially asked to complete a questionnaire about themselves and their trip. The trade-off task was then explained to them. Respondents rated each of the five trade-off cards (representing alternative trips) on five scales placed side-by-side on the back of the questionnaire they had just completed (appendix twelve).

Respondents seemed to have more difficulty with the photograph based trade-off. The textual scale trade-off, surprisingly, appeared to work more effectively.

6.5.2. Order Of Card Presentation:

It was found (in the pre-tests and pilot) that it took some time for the respondents to get going; this was especially the case where respondents were initially presented with one of the more difficult trade-off options.

As respondents were only considering five cards, it was thought that another easy card could be added to the beginning of each test - that would not be used in the valuation of ride. This card could have been used to illustrate the exercise without the interviewer having to worry about leading the respondent while explaining the task. But it was clear that the trade-off procedure took too long; respondents in the pre-tests and pilot tended to lose interest after five cards. Adding another card would have further increased the task.

It was therefore decided to always present respondents with the simplest of the five cards (Red) first, where all attributes were worse, risking a small order effect. The other four cards were still presented in random order.

The presentation order for each respondent was established using random numbers and the algorithm below. This order was written on to each questionnaire, before the interviews took place to save time. Once presented the cards were accessible to the respondent throughout the exercise; this should not have presented any difficulties as ranking approaches involve the respondent having access to all cards simultaneously. Respondents were allowed to alter their initial ratings if they wished to, later in the exercise. Few respondents volunteered changes; though a number were encouraged to consider it if they had appeared inconsistent.

FIGURE 8.4: CARD ORDERING ALGORITHM.

DRAW 1:

0.000-0.250: Blue, 0.251-0.500: Green,
0.501-0.750: Yellow, 0.751-1.000: White

DRAW 2:

0.000-0.333: Colour 1, 0.334-0.666: Colour 2,
0.667-1.000: Colour 3

DRAW 3:

0.000-0.500: Colour 1, 0.501-1.000: Colour 2

The trade-off cards were colour coded - not numbered or lettered so as not to give any indication of order. The colour was written on each card to avoid any psychological effects from colour preferences. The cards were coded as: Red (R), Blue (B), Green (G), Yellow (Y) and White (W). These colour codes are shown in table 8.2 (section 6.4). The order of the attributes was varied across the trade-off cards, as this has been found to affect results (section 6.2).

It was found that the process could be speeded up considerably by conducting group interviews. This was only possible where passengers were sitting close together, which could mean that respondents interfered with each others results. With group interviews each respondent assessed the cards in the same order - which may have caused some order effects. However group interviewing was not conducted on a large scale - despite being much more efficient.

6.5.3. Trade-Off Card Design:

As the attributes were in different places on each card, some respondents had difficulty referring to them. To rectify this, recognisable images were printed for each attribute. A clock was chosen for journey time and a piggy bank for fare. Initially a, "£" sign was to be chosen for fare, but this would have inferred larger changes in fare than were actually.

Some people had difficulty in establishing that each card represented an alternative trip (thinking that they had to trade-off attributes within cards). As a result of this, the titles of the cards were changed from, "COLOUR CARD" to, "COLOUR TRIP".

With the textual scale based trade-off, both ride levels were presented on one graphic scale making it easier for the respondent to see the difference between them (appendix fourteen). One difficulty with this approach was that, if the representation of today's trip on the ride scale did not match the respondents beliefs, he may have been confused. If this had happened on a wide scale this approach was unlikely to be successful.

With the photograph based trade-off (appendix fifteen) ride was described as being, "The same as today" or, "Like the train shown in the photograph".

6.5.4. Error Trapping:

With minimal supervision approximately 20-30% of respondents made significant errors in the trade-off process. However, with the improved supervision techniques that evolved during the interviews, wastage was reduced to perhaps 10%.

Some of the trade-offs done incorrectly in the early interviews were erased (they were done on pencil) and used again to maximise the valid sample size.

As the exercise progressed the interviewer became more skilled in spotting mistakes when they happened. Respondents could then be talked through apparent misjudgements to see what they really meant. It was clear that for the trade-off to be done effectively, respondents should be referred back to their previous assessments; as not all respondents were able to scale consistently unless this was the case. These processes significantly improved the results.

Most respondents managed the task eventually, generally older respondents appeared to have most difficulty. Some respondents who initially appeared too young to be interviewed gave the best responses. If it was clear that a respondent was not going to manage the task (very rare) they were asked some general questions about their journey and the interview was gently terminated.

The Yellow and White cards caused the most difficulty, as the same attributes changed (in different directions) on both cards. Some passengers failed to distinguish between these cards.

6.5.5. Conclusions:

Although the trade-offs were successful, the experience gained during their application means that the results would be significantly better if the exercise were repeated. However in practical terms doing a trade-off (even a very simple one like this) is perhaps pushing the limits of what is possible with unpaid respondents in such a difficult environment.

6.6. Coding:

Data from both trade-offs were entered into spreadsheets (appendix sixteen). The percentages of fare and journey time that the standardised trade-offs represented, for each person, were then calculated. Where respondents had failed to indicate their fare the trade-off data could not be used, as everything was measured in terms of fare in the analysis (two of each trade-off approach were lost because of this).

Irrational ratings were corrected as far as possible. This was sometimes necessary with the earlier trade-offs done before the interviewer had established an effective error trapping procedure (section 6.5.4). Alterations were only made that would not require any assumptions about a respondent's relative value of any attribute.

The most frequent problem was inconsistent scaling. The Red trip (always assessed first) was clearly the worst alternative and so a respondent's Red scores should have been the most negative. But some respondent's scaling became more severe after the Red assessment and so a later alternative could be assessed more severely. In such cases the analysis would often produce negative values for attributes.

The solution was to increase the Red score to match the later alternative's score. As long as the Red score was not placed any lower than this no judgement was being made for the respondent. Of the seventy-six valid trade-offs produced by the analysis, four of the (thirty-eight valid) photograph and nine of the (thirty-eight valid) textual scale based trade-offs had to be altered in this way. More of the textual scale trade-offs were altered, as they were generally the first to be administered.

The commonest demonstration of inconsistent scaling, was respondents giving both White and Yellow cards similar ratings (section 6.5.4). Where this occurred the trade-off data was not used in the analysis, as these ratings could not be adjusted without making considerable assumptions about the respondent's tastes (six textual scale and seven photograph based trade-offs were lost because of this).

6.7. Analysis:

6.7.1. General Procedure:

Multiple Regression was used to estimate the value of each attribute (section 5.4). The trip ratings formed the dependent variable, while the levels (-1, 0, +1) of the three attributes (ride, fare and journey time) formed the independent variables. The relative sizes of the attributes' coefficients produced by the regression indicated their effect on the trip rating (utility). The ride and journey time coefficients were then related to the % changes considered by each respondent (figure 8.5) so that an estimate could be made of each attributes value in terms of fare.

FIGURE 8.5: EXTRACTION OF RIDE AND JOURNEY TIME VALUES.

$$\text{Ride Value (\% of fare)} = {}^{RCo}/_{FCo} \times \%ChFare$$

$$\text{Journey Time Value (\% of fare)} = \frac{{}^{JTCo}/_{\%ChJTime}}{{}^{FCo}/_{\%ChFare}} \times 100$$

RCo	=	Ride Coefficient.
FCo	=	Fare Coefficient.
%ChFare	=	Percentage Change in Fare considered by Respondent
%ChJTime	=	Percentage Change in Journey Time considered by Respondent.
JTCo	=	Journey Time Coefficient.

Two methods of calibration were used. Firstly, one assessment from each respondent (plus eight null assessments, to make the data set more orthogonal (section 6.4)) was regressed to obtain a cross-sectional valuation. Population values for each attribute were estimated using the standard errors of the regression coefficients.

Repeated measures calibrations were also done. The five assessments produced by each respondent (plus one null assessment) were regressed to estimate each person's attribute values. The mean of these estimates was taken and the standard error calculated, to provide estimates of the population's values.

6.7.2. Repeated Measures V' Cross-Sectional Calibration:

A cross-sectional calibration of this data uses only one rating from each respondent; we would therefore expect these results to be less valid than any produced using a repeated measure calibration - which would be based on a all the respondents' ratings.

To use multiple regression, we must have a sample size of at least one more than the number of independent variables. In practical terms the sample size should be at least four times the number of independent variables (Groebner and Shannon (1985)). This latter criterion could not be met while using repeated measures (six assessments and three independent variables). However by combining a large number of individuals' results, the likelihood of extreme values should be greatly reduced.

One problem with a cross-sectional calibration, is interpersonal variation in scaling. Each respondent may perceive a different level of utility from the same position on a rating scale. This will cause the valuation estimates from a cross-sectional calibration to be less precise (section 5.3) than with repeated measures.

One aim of this study was to be able to produce different ride values according to personal characteristics - to allow ride values to be generated for different lines in the future. This is much easier with an individually calibrated model, as groups of certain types of individuals ride values can be aggregated. If a cross-sectional calibration is used, isolating certain types of individuals, means that the design is no longer orthogonal leading to less predictive accuracy in the model.

Finally as the number of valid respondents was not in multiples of five, a cross-sectional calibration of this data would not be orthogonal (sections 5.4.1. and 5.5) making the regression estimates less precise.

It is clear from this discussion that the repeated measures calibrations are likely to produce the most useful results.

7. CONCLUSIONS:

This chapter has seen the detailed development of three methods for establishing values of a change in engineering ride. In the following chapter the results of the implementation of these methods are described and discussed.

Chapter Nine Valuation of Ride - Results

1. INTRODUCTION:

In this chapter the final valuation results are presented and discussed. The results from all three approaches are considered in the light of previous research and compared with each other. An attempt is made to produce some form of model based on these results. Finally there is an effort to produce a single figure for the value of a change in engineering ride quality.

2. COMPARABILITY OF RESULTS:

The results from all three techniques are reported in terms of % of fare, so that they may be compared with each other and the previous results produced by M.V.A. It should be remembered that the weak relationship between M.V.A's ride values and engineering measures means that comparisons must be treated cautiously.

All three valuation approaches were based on the same two trains, old D.M.U.s and replacement Sprinters. These broadly represent the two levels of engineering ride that were valued by respondents (chapter eight). As mentioned earlier (chapter seven) there were a number of problems that limited the number of engineering measurements taken after the introduction of the Sprinters. This means that engineering measurements for both the old D.M.U.s and Sprinters are only available for the Castle Cary to Yeovil Pen Mill section. The relationship between ride values and engineering measures cannot therefore be specified as closely as was hoped.

The locations on the textual ride scale are precisely related to engineering ride, as these were generated from assessments of ride on the Castle Cary to Yeovil Pen Mill section only.

Respondents doing the other two tasks (photograph trade-off and stock replacement) were valuing the difference in ride between the Sprinters and the old D.M.U.s for the whole of their trip. As each of these approaches was based on a different sample, there is likely to be some difference in the engineering measures valued with each approach. Very few passengers were making trips from Castle Cary to Yeovil Pen Mill.

Castle Cary to Yeovil Pen Mill was one of the poorer sections of track. As differences in engineering ride between stock may vary according to track quality (chapter two) this could mean that most respondents were assessing different changes in engineering ride than our measurements indicate.

This unrepresentative measurement of engineering ride does not substantially reduce the value of the findings, as the ride valuations are nevertheless directly related to engineering ride (unlike M.V.A's findings). The main aim of this thesis was not to concentrate on the taking of engineering measurements. If required further engineering measurements could be taken to provide more representative figures.

It should be noted from this discussion that the results of the three valuation approaches are not strictly comparable, as they could all be valuing different changes in engineering ride. However it was shown in chapter seven that the levels of engineering ride associated with different overall trips varied little (as good and bad sections even each other out). It is thus believed that the results of the photograph based trade-off and the stock replacement approaches are compatible. Comparisons of these approaches with the textual scale trade-off must be treated with more caution.

3. CHANGE IN STOCK APPROACH:

3.1. Results:

One hundred and four questionnaires were distributed to eligible passengers, who seemed to have little difficulty with the exercise. Only six of these respondents stated that they would not have made today's trip if it had been associated with the lower ride quality of the old D.M.U. This represents a loss of 5.77% of passengers.

To estimate the proportion of passengers that would be lost, if we had interviewed the whole population of rail users, we need to look at the standard error of the population proportion (Harper (1982)):

Standard Error = $\sqrt{(^{p*q}/_{n})}$

We have p = 0.0577, q = 0.9433 and n = 104, so the equation is:

$$\sqrt{((^{0.0577*0.9433)}/_{104})} = 0.023$$

Applying this standard error to the proportion of passengers lost in the sample, produces a population estimate of: 1.17% - 10.37%.

To be able to compare this result to those of other approaches (and M.V.A's) we need to convert the 5.77% estimated change in patronage into a value in terms of fare. From previous work (British Railways Board (June 1986)) we know that a 5.77% reduction in patronage could be caused by a 5.77% increase in fare (elasticity of -1) we can therefore say that this change in ride is equal to a 5.77% change in fare. The population's value of this change in ride thus lies between 1.17 and 10.37% of fare.

3.2. Validity of Result:

The broad band of the population estimate illustrates one of the major problems with this technique. As so few respondents indicated a change, just one more/less person deciding not to travel would have had a significant effect on the results. The population band would be most effectively reduced, by increasing the sample size. As this technique is so straightforward, this would not be difficult.

A further difficulty with this approach is, that it is obvious to respondents what their answers are meant to show. It is therefore easy for a respondent who wants an improvement to the service to exaggerate their response to the ride change. As this approach was designed for self completion, we have little indication of what went through the minds of the respondents while they completed the task. However from speaking to respondents before and after the exercise and reading their comments on the questionnaires, it appears that the technique has been successful. It was interesting to note that a number of people stated that they would still make today's trip with the rougher ride, but would not like it. This implies that any, "Protest vote" has been contained.

Although this technique was designed to minimise contamination, there is bound to be some element of this in the results. All respondents will to some extent have been affected by other differences between the two trains. One of the questionnaires was rejected as the respondent's comments showed indications of this.

Perhaps the best indication of the effectiveness of the technique is the value of ride that it produced. If the biases just discussed had been important we would have expected a very high value of ride; but this approach produced the lowest value of ride.

4. TRADE-OFF APPROACHES:

4.1. General:

After coding and analysis the final valid sample sizes for both trade-offs emerged as: thirty-eight photograph (out of fifty) and thirty-eight textual scale (out of forty-nine) based trade-offs.

4.2. Repeated Measures Results:

The results of the repeated measure calibration (for detail see appendix sixteen) were:

PHOTOGRAPH TRADE-OFF.

Ride: 8.29% fare - Range @ 95%: 4.55-12.03% fare.
Journey time: 59.8% fare - Range @ 95%: 35.44-84.16% fare.

TEXTUAL SCALE TRADE-OFF.

Ride: 8.47% fare - Range @ 95%: 3.53-13.29% fare.
Journey time: 54.62% fare - Range @ 95%: 23.64-85.6% fare.

The population estimates were produced using:

$$\text{Standard Error} = \text{Standard deviation}/\sqrt{n}$$

as recommended in (Groebner and Shannon (1985), Harper (1982)).

4.3. Cross-Sectional Results:

Although valuations were extracted using a cross-sectional calibration these are not presented below. For the reasons discussed previously (chapter eight) it was found that the results of this calibration were unreliable. The results varied considerably with the selection of individuals' ratings. For example, in three different cross-sectional calibrations (using the photograph based trade-off) ride values varied between: 4.06 - 8.29% of fare and journey time between: 32.43 - 41.76% of fare. All cross-sectional estimates were characterised by greater standard errors than the repeated measures results. This meant that less precise estimates of the population's values of ride and journey time could be made.

This result does not necessarily mean that a cross-sectional approach is inferior. It would be possible to design a trade-off specifically for cross-sectional application, where each person rates only one alternative trip using a self-completion approach (as was originally intended in this research). This would make a large sample practical, offsetting some of the disadvantages previously discussed (chapter eight). However for this research the risk of pursuing a solely cross-sectional approach was considered too great.

4.4. Validity of Results:

The results of the cross-sectional calibrations are not considered reliable enough to attach any significance to. However the repeated measures results are more significant. Even so, the trade-offs exhibited a wide range of valuation estimates for the population, a larger sample would have reduced this.

The two ride values extracted by the trade-off approaches are very similar (and not statistically different @ 95%). This implies that there was little difference in the changes in engineering ride valued by each group (section 2). The photograph based approach produced a value of: 8.29% of fare and the textual scale approach: 8.47% of fare. Both ride values are significantly (@ 95%) above the 2.17% (rough estimate - chapter eight, section 6.2) found by M.V.A. on InterCity services.

One effective way of establishing the validity of the ride values extracted from the trade-offs is to look at the journey time values produced by the same process and see how these compare with previous research. One difficulty with this is that, although considerable research has been done on the value of time, there is no specific Provincial journey time value. The most applicable value is that for InterCity services (British Railways Board (1986)) which is 40% of fare (an elasticity of -0.4).

The journey time values extracted from the trade-off are greater (though not significantly) than would be expected from the InterCity value (59.8% of fare for the photograph based trade-off and 54.62% of fare for the textual scale based trade-off). The difference between the journey time values extracted by each trade-off is not statistically significant (@ 95%). An obvious explanation for the trade-offs' higher values is that trains on InterCity routes are generally faster and thus more competitive. The route on which the trade-offs were done, was particularly uncompetitive and so one would expect journey time to be more highly valued by passengers. A number of comments were made to this effect during the interviews.

The population estimates of the textual scale approach are marginally less precise than those of the photograph based technique. The difference in precision of the ride values implies that the textual scale was a slightly less precise way of referring to levels of ride quality. However the journey time value, generated with the textual scale approach, was also less precise. Journey time values were produced in the same way with both trade-offs, suggesting that the results of the photograph based trade-off are generally more precise. The photograph based trade-offs tended to be done towards the end of the interview period. As the quality of the trade-off data improved during the interviews, the differences in precision are probably a reflection of this (chapter eight).

It is important to realise that people may respond differently to positive and negative changes of attributes. There was some some indication of this during the the research. A worsening of an attribute was nearly always associated with a lower rating. An improvement in an attribute did not always result in a higher rating. The trade-off values presented in this research are averages and may be higher if a reduction is being considered, or lower with an improvement.

The plausible journey time values and the fact that the results of both trade-offs are so close, suggests that the results are valid. However care should be taken in their implementation, because of the wide range of population estimates, resulting from small sample sizes.

5. MODELLING RIDE VALUES:

An attempt was made to develop a model, that would estimate ride values for lines with specified characteristics. This was done with the data from both trade-off approaches. But the relationships between ride values and personal/trip characteristics were so weak (table 9.1) that no significant regression models could be constructed. The coefficients in all the models developed were insignificant, while the maximum R2 value achieved was: 0.07. The main cause of this was the small sample sizes.

TABLE 9.1: CORRELATIONS BETWEEN RIDE VALUES AND PERSONAL/TRIP CHARACTERISTICS.

PHOTOGRAPH TRADE-OFF:

Ticket Price: - 0.03, Journey Time: - 0.00,
Passenger's Age: 0.10, Passenger's Sex: 0.12,
Regularity of Rail Use: - 0.03

TEXTUAL SCALE TRADE-OFF:

Ticket Price: - 0.15, Journey Time: - 0.24,
Passenger's Age: 0.01, Passenger's Sex: 0.09,
Regularity of Rail Use: 0.04

The relationship between ride values and trip type was also examined, using an ANOVA. First the variances of each trip type had to be compared, as a standard ANOVA works on the assumption of each group having equal variances. It was found that there were significant differences in the variances of the trip types (@ 99%) and A Kruskall-Wallis Non-parametric ANOVA was used. A null hypothesis was generated that, "There is no significant difference between the ride values of passengers on different trip types". This was done for both sets of trade-off data.

PHOTOGRAPH TRADE-OFF:

Kruskall-Wallis Statistic (corrected for ties) = 3.25.

Not significant @ 80% with 4 d.f.

TEXTUAL SCALE TRADE-OFF:

Kruskall-Wallis Statistic (corrected for ties) = 3.92.

Not significant @ 80% with 4 d.f.

Again no significant differences were found, probably as a result of the small sample sizes.

6. SUMMARY AND CONCLUSIONS:

Values of a change in engineering ride have been generated by each of the three approaches developed in this research. The best estimate of the change in engineering ride that has been valued by respondents, in all three approaches, is based on measurements taken between Castle Cary and Yeovil Pen Mill (section 2). The lateral difference is 15 (23 - 8) and vertical difference 25 (40 - 15). These measurements are all in I.S.O. weighted m/s2 R.M.S. There is no agreed way of combining lateral and vertical accelerations, so these have to be reported separately. The inability to provide a reliable single measurement of a change in ride quality, makes it almost impossible to produce an elasticity for ride.

The lowest ride value, 5.77% of fare, was found using the stock replacement approach. The photograph based trade-off produced a value of 8.29% of fare and the textual scale based trade-off 8.47% of fare. All these estimates are associated with large standard errors, as a result of small sample sizes and so there is no significant difference between them. The fact that the three values are so close does to some extent validate the results.

The possible policy implications of these results, means that it is important to compare the values produced in this research with those of earlier work. But as stated earlier (chapter two) the only previous values of a change in ride quality (produced by M.V.A) are not unambiguously related to engineering measures. Comparisons have thus been made on the basis of the assumptions made in chapter eight (section 6.2). Considerable care must be taken in the interpretation of such comparisons, as the loose relationship between M.V.A's results and engineering measures of ride could easily account for a large part of any discrepancy.

The ride values produced by both the trade-offs are significantly higher (@ 95%) than the approximate values produced by M.V.A. for InterCity. The value produced by the stock replacement approach is also higher (though not significantly). Some of this difference is probably a results from interference in the ride ratings. As the M.V.A. approach was not related to any particular stock, we would not expect their results to so affected. However the textual scale trade-off was less closely related to stock types and this produced the greatest value of ride. M.V.A's estimates were produced on InterCity services where ride quality is generally higher and perhaps therefore less of an issue.

In chapters six and seven attempts were made to estimate the size of interference effects, by comparing passengers assessments of ride quality with their perceptions of the levels of other attributes. If this earlier work is valid in this context, interference could be responsible for up to 42% of the ride values extracted in this chapter. In such a case the average ride value, produced by this thesis, of approximately 7% of fare would be reduced to 4-5%.

Journey time values were also extracted from the trade-offs. Although these values are higher (not significantly) than those found in previous InterCity research, they do appear plausible - further validating the ride results.

Attempts to relate ride values to personal and trip characteristics proved unsuccessful, because of the small sample sizes.

It should be remembered that all stated preference research (in dealing with people) is vulnerable to mistakes and even deliberate misinformation. There are many such factors that may cause the results of this research to be inaccurate. A number of these are discussed in chapter five when the principles of scaling are discussed. For example, the presence of the interviewer may have caused respondents to over/understate their reactions.

Overall the best estimate of the value of the difference in ride, between a Sprinter and an old D.M.U. would between 4-7% of fare. If a single figure has to be chosen, a pessimistic value of 5% of fare is suggested. It therefore seems that M.V.A's vague ride values currently used by Provincial may be too low (and definitely capable of too broad an interpretation). As all three of this study's approaches have produced similar results that are directly related to engineering ride measures, these values appear more valid in a rural Provincial context. It also appears that the previous values of journey time used by Provincial, may be too low.

This ride value can be directly related to levels of passenger demand. The change in engineering ride between and old D.M.U. and a Sprinter has 5% of the effect of a fares change. Any change in fares is thought to have a unitary (100%) effect on patronage, this change in ride would thus be expected to have a 5% effect on patronage.

Chapter Ten Summary and Conclusions

1. INTRODUCTION:

This research has examined the way in which intangible product attributes and in particular ride quality affects the demand for rail travel.

In an attempt to improve the financial position of rural routes, British Rail Provincial have decided to try and optimise the relationship between track maintenance and revenue. Provincial knew the costs involved in providing a certain level of engineering ride, but they did not know about the relationship between engineering ride and the demand for the service. There was thus a need for further information about this relationship.

One of the main aims of the study was thus, to find a method that could establish the financial value individuals attached to changes in intangible attributes like ride. This has been achieved.

2. METHOD USED:

2.1. General:

The work has been done using stated preference techniques. A series of interviews were carried out which got progressively more structured and specific as the topic developed.

Initially all the possible methods of producing a value for a change in engineering ride were considered (chapter three). Some of these could be immediately eliminated because of knowledge from previous research (chapter two). Depth interviews were carried out (chapter four) allowing an investigation into the general issues involved with intangible attributes. This process allowed a form of language to be developed that was comprehensible to the public.

Once this information had been processed the remaining methods of generating a ride value were developed in more detail (chapter five). A ride scale that could be comprehended by the public was developed at this point. A series of interviews were then conducted in which the relationships between intangible attributes were studied (chapters six and seven). Some of the remaining approaches were eliminated on the basis of information obtained during this stage.

Finally the three remaining valuation approaches were implemented (chapter eight). All three approaches were broadly compatible, so that their results could be compared. Two of the approaches involved a trade-off exercise, while the third asked passengers to remake their travel decision on the basis of a reduction in ride quality. Results were produced in a form that allowed comparisons with the only previous study that had attempted to estimate a value for a change in ride quality.

2.2. Applicability of the Method:

The valuation methods developed during this research could be applied to other intangible attributes in other areas.

Intangible attributes are characterised by a scale of measurement that is not readily comprehended by the layman. Thus for a change in the level of an intangible attribute to be valued, some alternative form of measurement has to be developed that is meaningful to the public. This scale should be related to technical measurements, so it is clear to the researcher what changes are being valued.

Two main forms of alternative scaling were developed during the thesis. Levels of attributes were referred to in terms of the levels associated with a certain product, for example the ride on the Sprinter. A textual scale was also developed (chapter five) using the language and ratings of the public. This second scale consisted of a straight line with anchors at each end, for example, "Smooth ride - don't notice the train is moving". Levels of the attribute could be represented by a point along the line.

Both these forms of scaling could be used in other areas. Though the textual scale would need to have new anchors developed, based on a typical form of language obtained from depth interviews. As an example consider the comfort of seating. We could refer respondents to the levels of seating comfort on various trains, cars etc. Alternatively we could generate a textual scale using anchors such as, "Very comfortable, even after a long journey".

A second problem when establishing a value for changes in an intangible attribute is the difficulty of isolating these changes, from changes in other attributes. Some attempt has been made in this study to isolate and contain such Interference effects. Although ride may be particularly susceptible to such effects because of its close relationship with noise; account should still be taken of Interference effects if these methods are to be applied elsewhere.

Once a form of scaling is established, an attempt could be made to value any intangible attribute using similar techniques to those developed in this thesis. One could ask current users of a product whether they would continue to use it, if an attribute level was worsened. Alternatively respondents could participate in some form of trade-off exercise, in which the values they attach to various attributes are revealed.

One interesting issue to arise from the application of the three valuation approaches is that the simplest approach produced similar (slightly lower) results to the more complex trade-off approaches. Fears that asking respondents directly for their response to a change in ride may produce exaggerated results, appear unfounded in this case. If similar results occurred with larger samples it could have important implications for the design of future research.

3. GENERAL FINDINGS OF THE RESEARCH:

The best phrase found to describe ride quality was, "Smoothness of ride" (chapter four). This was used successfully throughout the research.

Relationships between ride values and personal/trip characteristics could not be established because of the small size of the valuation samples (chapter eight). Relationships were established between the severity of ride ratings and personal/trip characteristics (chapters six and seven).

Generally it was found that frequent, work type trips were the most critical of ride and that men may be slightly more critical. Older people and those with less experience of standing were slightly less critical. From this one would expect people on commuter type services to be the most critical of ride quality.

There was an indication that assessments of intangible attributes interfere with each other. Noise could be a strong contaminant in ratings of ride quality, this relationship was expected from previous work on ride quality. Other attributes were also found to interfere with ride assessments, but their effects were small compared to noise. Some indication was provided of the size of these interference effects, but the limited data on which this estimate is based should be stressed. In the worst case as much as 42% of ride ratings could be a result of such interference effects.

Passengers were able to distinguish between the engineering ride of trains of a different generation (for example, old D.M.U's and Sprinters) but their ability to distinguish between the ride of various track sections is questionable. People were found to give more severe ratings (approximately 20% higher) if they knew in advance what they would be assessing (hyper-sensitivity).

Individuals' asked to assess the previous section of line (after they had experienced it) produced assessments that were just as reliable, as those produced while the section was being experienced.

Most passengers were found to have made their decision to travel at home. Each respondent produced an assessment of the importance of a ride improvement on-train and at-home. No significant difference was found between these two assessments, indicating little or no environmental effect (though there were weaknesses in the method).

It was found that people were able to fairly accurately perceive levels of engineering ride from positions on the textual ride scale developed. This has important implications for the study of other intangible attributes.

Values of journey time were produced by the research, no such value has been produced for Provincial services in the past. Previously Provincial have based the development of their services on the InterCity value of time (40% of fare). Both values produced by this research are higher (but not significantly @ 95%) than the InterCity values (average 57% of fare). There is thus an indication that journey time on Provincial services may be more important than was thought. As speeds on Provincial services are lower than those of InterCity this result is plausible.

4. THE RIDE VALUES:

Financial values for a change in engineering ride were produced using three different approaches. All the approaches produced ride values that were close: 5.77% of fare, 8.29% of fare and 8.47% of fare. There were wide standard errors around these estimates. Even so the latter two values were significantly different from the value found in previous research of approximately 2.17% of fare (though the latter figure is not closely related to engineering measures and so may not be strictly comparable).

The existence of interference effects means that these ride values are probably exaggerated. It is suggested that the best single estimate of the value of the change in engineering ride considered (lateral 15, vertical 25, I.S.O. weighted m/s2 R.M.S) is 5% of fare.

Care should be taken in the implementation of this figure, as it is based on a small sample of approximately two-hundred. It is also likely that increases and reductions in attribute levels are treated differently. The figure produced in this thesis (5%) is based on a reduction in engineering ride. M.V.A's (May (1986)) approximate figure of 2.17% is based on an improvement in ride quality. It was found during the valuation exercises that respondents tended to consider reductions in attribute levels more severely than improvements. This may partly explain the difference between the ride values produced in this research and those of M.V.A. If a value is needed for an improvement in ride quality, 5% could therefore be too high.

5. APPLICATION OF RESULTS:

If the values produced by this research were to be used in any financial context, it would probably be wise to expand the size of the sample. This would reduce the standard errors of the estimates and establish more precise values.

Some of the findings suggest that certain groups of people and types of trip are more critical of ride quality. It may be that such people/trips have different values of ride. This research was unable to establish a relationship between ride values and person/trip characteristics because of the small sample size. If ride values are to be applied to any lines that are significantly different from the one where this research was done (Bristol to Weymouth) it would be wise to study these relationships in more detail.

It would be difficult to choose one particular technique for further use, as they all appeared to work equally well. If the exercise was to be done with very limited resources a self-completion technique would be most appropriate: either the stock replacement approach, or maybe some form of cross-sectional trade-off. If considerably more resources were available, it may be possible to modify the techniques developed in this research and alter the suspensions of trains or coaches.

The survey location would also be important in deciding which technique to use. If stock with different levels of engineering ride was not available for respondents to assess, the textual scale trade-off would have to be used.

6. IMPLICATIONS OF THE RESULTS:

The main need for a financial valuation of ride was so that Provincial could optimise track standards. Engineering ride has recently been significantly improved on most routes by the introduction of new Sprinter trains. It may now be possible to reduce track maintenance and the costs associated with it, without loosing many passengers. Costs would thus be reduced by a greater amount than revenue.

As the ride value produced by this thesis may be higher than previously thought, a reduction in ride quality may have a greater effect on patronage than was expected. This would make the idea of a reduction in track standards less attractive, reducing the scope for cost savings on Provincial services.

As a reduction in ride quality may have a more severe effect on patronage than an increase, it may have been better to reduce track quality when the Sprinters were introduced. Passengers would then have experienced a small improvement in engineering ride, rather than a big improvement and then a reduction (section 4).

7. RECOMMENDATIONS FOR FURTHER WORK

The small sample sizes used in the valuation exercises have already been mentioned. Further work is therefore necessary to provide more robust estimates of a ride value based on larger samples.

An investigation also needs to be done into how ride values vary with personal and trip characteristics. Ideally some form of model would be produced to generate ride values for various types of service. An attempt was made to develop such a model in this thesis, but the small sample sizes meant that no significant relationships emerged.

It was mentioned in the previous section that a reduction in ride quality may be judged more severely than an improvement. This phenomena suggests that if a service is improved and then reduced to its original level, there may be a net loss of passengers. It may be that once passengers have experienced a higher level of quality their expectations are raised and they become less tolerant of the original level. Further research may provide some interesting insights into this issue.

References

Aaker D.A. & Day G.S.	Marketing Research, 3rd Edition. Wiley, New York (1986).
Abbott J.	Resurgence in the Valleys. Modern Railways, Pages 138-142 (March 1986).
Abbott Stan.	To Kill a Railway. Leading Edge Press, Hawes (February 1986).
A.B. Group Ltd.	A Comparative Attitude Survey... Response to Class 141 D.M.U's. Consultants Report (June 1984).
Ackerman L.J, Brinberg D. & Jaccard J.	Assessing Attribute Importance: A Comparison of Six Methods. Journal of Consumer Research, Pages 463-468 (March 1986).
Advertising Principles Ltd. 1986).	A Comparative Attitude Survey on the Response to Class 143 D.M.U's. Consultants Report (November
Allan Ian.	Sectional Maps of Britain's Railways, Third Edition. Ian Allan, Shepperton (1989).
Apperson J.	The Radio Electronic Token Block Signalling System.... Rail International, Pages 45-49 (February 1987).
Applied Psychology Unit.	Conducting a Survey. Applied Psychology Unit, Cranfield - Student Notes (1988).
Aspinall D.T.	Vehicle Riding Comfort: A Comparison of Subjective and Objective Measures. M.I.R.A. Report 1960/4, Lindley (1960).
Aspinall D.T. & Oliver R.J. (1964).	Vehicle Riding Comfort: Correlation of Subjective and Physical Measures. M.I.R.A. Report 1964/10, Lindley
Association of County Councils & British Railways Board.	Review of Rural Railways. A.C.C. and British Rail (1984).
Baddeley A.	Your Memory - A Users Guide. Penguin, New York (1982).
Baker S.K.	Rail Atlas of Britain and Ireland, Fourth Edition. Oxford Publishing Co., Poole (1984).
Barnard P.O, Hensher D.A & Truong T.P.	The Role of Stated Preference Methods in Studies of Travel Choice. Journal of Transport Economics and Policy, Pages 45-58 (January 1988).
Beazley D, Lunn T. & Westwood D.	The Trade-off Model and it's Extensions. Journal of the Market Research Society, Pages 227-241, Number 3 (1974).

Beeseley & Kettle.	Improving Railway Financial Performance. Gower, Aldershot (1984).
Bell, Bowen & Fawcett.	The Business of Transport. Macdonald and Evans, Plymouth (1984).
Ben-Akiva M. & Lerman S.R.	Discrete Choice Analysis. M.I.T. Press, Cambridge Mass. (1985).
Benwell M, Black I.G, Hannan T.R.D & Seaton R.A.F.	Train Service Reliability on British Rail InterCity Services. Centre for Transport Studies, Cranfield (1985).
Beresford P. & Welsh E.	Trouble on Tracks for B.R. as Lightweight Trains Crack Up, Sunday Times (1986).
Best R. & McCullough J.	Conjoint Measurement: Temporal Stability and Structural Reliability. Journal of Marketing Research Special Section - Measurement and Market Research, Pages 26-31 (February 1979).
Bither S.W. & Klein N.M.	An Investigation of Utility Directed Cut-off Selection. Journal of Consumer Research, Pages 240-256 (September 1987).
Bly P.H. & Webster F.V.	The Demand for Public Transport, Transport Reviews. Pages 23-46 (April 1981).
Bly P.H. & Webster F.V.	The Demand for Public Transport, Transport Reviews. Pages 323-351 (January 1982).
Bonavia M.R.	The Birth of British Rail. George Allen and Unwin Ltd., London (1979).
Bonavia M.R.	Twilight of British Rail. David and Charles, Newton Abbott (1985).
Bradley M.	Realism and Adaptation in Designing Hypothetical Travel Choice Concepts. Journal of Transport Economics and Policy, Pages 121-137 (January 1988).
Bradley U. (etd.)	Applied Marketing and Social Research, First Edition. Van Nostrand Reinhold Co., New York (1982).
British Rail Education Service.	An Introduction to British Rail. Public Relations Hand-out (1985).
British Rail, Provincial.	Reactions to New Rolling Stock. Internal Report (1982).
British Rail, Provincial.	Perfect Escapes by Rail and Ferry. Promotional Leaflet (1985).
British Rail,	What is British Rail Provincial?

Provincial.	Promotional Leaflet (April 1986).
British Rail, Provincial.	Sprint to the Hills. Promotional Leaflet (1987).
British Rail, Provincial.	Britain's Scenic Rail Journeys. Promotional Leaflet (1987).
British Rail, Railway Technical Centre.	Network South East Ride Quality Study - Preliminary findings. B.R. Internal Report, Derby (1987)
British Rail, Railway Technical Centre.	Evaluation of the Effects of Easing the Allowable Cant Deficiency - W.C.M.L. B.R. Internal Report, Derby (date unknown).
British Railways Board.	The Reshaping of British Railways. H.M.S.O., London (1963).
British Railways Board.	A Comparative Study of European Railway Performance. British Railways Board, London (December 1979).
British Railways Board	Corporate Plan 1985. British Railways Board, London (1985).
British Railways Board	Sprinter and Pacer. Promotional Leaflet, British Railways Board, London (1985).
British Railways Board.	Sprinter and Pacer Diesel Multiple Units - Technical Description. Public Relations Hand-out, British Railways Board, London (1986).
British Railways Board	New Look at B.R's Provincial Sector. Public Relations Hand-out, British Railways Board, London (1986).
British Railways Board.	Passenger Demand Forecasting Handbook. British Railways Board, London (June 1986).
British Railways Board	Annual Report and Accounts. British Railways Board, London (1989).
British Standards.	British Standard 6841 - Measurement and Evaluation of Human Exposure to Whole-Body Mechanical Vibration and Repeated Shock. British Standards, London (1987).
Carroll J.D, Green P.E. & Goldberg S.M.	A General Approach to Design Optimisation via Conjoint Analysis. Journal of Marketing (Summer 1981).
Cattin D.R. & Wittink P.	Commercial Use of Conjoint Analysis - A Survey. Journal of Marketing, Pages 44-53 (Summer 1982).
Cattin D.R. & Wittink P.	Commercial Use of Conjoint Analysis: An Update. Journal of Marketing, Pages 91-96 (July 1989).

Central Transport Consultative Committee.	Annual Report. C.T.C.C., London (April 1979).
Central Transport Consultative Committee.	Rural Railways. C.T.C.C., London (September 1979).
Central Transport Consultative Committee.	Annual Report. C.T.C.C., London (June 1987).
Central Transport Consultative Committee.	Annual Report. C.T.C.C., London (June 1988).
Centre for Transport Studies.	Customer Reaction to the Introduction of Sprinter Units on the Cardiff to Weymouth Line. Centre for Transport Studies, Cranfield (1989).
Centre for Transport Studies,	Report on Crewe-Cardiff Line, M.Sc. Group Project. Centre for Transport Studies, Cranfield (1989).
Chappell T.A.	Passenger Comfort Tests of June 1983 - Planning and Conduct. British Rail Railway Technical Centre Internal Report, Derby (April 1984).
Chappell T.A.	Passenger Comfort Tests (A.P.T.) of April 1984 - Planning and Conduct. British Rail Railway Technical Centre Internal Report, Derby (February 1986).
Clayton G. & Rees J.H.	The Economic Problems of Rural Transport in Wales. Wales University Publications, Cardiff (1967).
Cohen L. & Holliday M.	Statistics for Social Scientists. Harper and Row, London (1982).
Cohen J.B. & Miniard P.W.	Isolating Attitudinal and Normative Influences in Behavioural Intentions Models, Journal of Marketing Research Special Section - Measurement and Market Research,
Pages	102-110 (February 1979).
Cooper B.K. (etd.)	British Rail Handbook. Ian Allan, Shepperton (1984).
Cooper D.R, Jordan P.G. & Young J.C.	Road Surface Irregularity and Vehicle Ride - Part 1. T.R.R.L. SR 341, Crowthorne (1978).
Cooper D.R. & Young J.C,	Road Surface Irregularity and Vehicle Ride - Part 2. T.R.R.L. SR 400, Crowthorne (1978).
Cooper D.R. & Young J.C.	Road Surface Irregularity and Vehicle Ride - Part 3. T.R.R.L. SR 560, Crowthorne (1980).

Copley G. et al	Station Investment: Priorities and Values. Energy Transportation and the Environment, Proceedings from the ESOMAR Conference, Montreux, Switzerland, Pages 21-45 (1987).
Dare A.	Radical Changes on Local Rail Services in the Netherlands. Eleventh Annual Seminar on Rural Public Transport P.C.L., London (1983).
David G,	B.R. faces Writ as Trains Fail. Sunday Times (January 1988).
Department of Transport.	The Cambrian Coast Line. H.M.S.O., London (1969).
Dorset County Council	Passenger Flows on the Dorchester-Yeovil Line. Unpublished Tables (1983).
Dorset County Council.	Public Transport Plan. D.C.C., Dorchester (1986).
Edwards A.L. W.H.	Multiple Regression and the Analysis of Variance and Covariance. Freeman, San Francisco (1979).
Fenwick I. Journal	A Users Guide to Conjoint Measurement in Marketing, European of Marketing, Pages 203-211, Part Two (1978).
Ford R.	The D.M.U. Replacement Story, Modern Railways. Pages 315-320 (June 1986).
Fowler D.	Financial Analysis of Rail... Transport (May 1979).
Fowkes T. & Wardman M.	The Design of Stated Preference Travel Choice Experiments with Special Reference to Inter-personal Taste Variations. Journal of Transport Economics and Policy, Pages 27-44 (January 1988).
Foxall G.R.	Consumer Choice. St. Martins Press, New York (1983).
Frederick C.O. Engineering	Optimised Track Standards, Paper to Provincial Sector/Civil Conference, British Rail (14th October 1987).
Freeman-Allen G. (etd.)	Janes World Railways 1985-6. Janes Publishing Co., London (1986).
Freeman-Allen G. Books	The Illustrated History of Railways in Britain. Marshall Cavendish Ltd., London (1985).
Friedman L.W. & Friedman H.H.	Comparison of Itemised v's Graphic Rating Scales.... Journal of the Market Research Society, Pages 285-289 (July 1986).

Gammell C.J. The Rural Railway. G.R.Q. Publications, London (1983).

Ginter J.L. & Miller K.E. An Investigation of Situational Variation in Brand Choice Behaviour and Attitude. Journal of Marketing Research Special Section - Measurement and Market Research, Pages 111-123 (February 1979).

Glover J. B.R. Diary 1958-67. Ian Allan, Shepperton (1987).

Glover J. B.R. Diary 1978-85. Ian Allan, Shepperton (1985).

Glover J. The Cost of the Social Railway. Modern Railways, Pages 52-53 (January 1988).

Government Statistical Service. United Kingdom in Figures. H.M.S.O., London (1989).

Green P.E & Rao V.R. Conjoint Measurement for Quantifying Judgmental Data. Journal of Marketing Research, Pages 355-363 (August 1971).

Green P.E. On the Design of Choice Experiments involving Multifactor Alternatives. Journal of Consumer Research, Pages 61-68 (September 1974).

Green P.E. & Srinivasan V. Conjoint Analysis in Consumer Research: Issues and Outlook. Journal of Consumer Research, Pages 103-123 (September 1978).

Groebner D & Shannon P. Business Statistics. Charles E. Merrill Pub. Co., Columbus Ohio (1985).

Guilford J.P. Psychometric Methods. Mc Graw-Hill, New York (1954).

Hamilton K. & Potter S. Losing Track. Routledge and Kegan Paul, London (1985).

Harborough P.R. Passenger Comfort During High Speed Curving - Summary Report. Railway Technical Centre, British Rail Internal Report (May 1986).

Harborough P.R. Passenger Comfort During High Speed Curving - Analysis and Conclusions. Railway Technical Centre, British Rail Internal Report (July 1986).

Haresnape B. British Rail Fleet Survey - 8 D.M.U's The First Generation. Ian Allan, Shepperton (1985).

Haresnape B. British Rail Fleet Survey - 9 D.M.U's The Second Generation and D.E.M.U's. Ian Allan, Shepperton (1986).

Harrell L. The Future Potential of Rural Railways. (B.A. dissertation) Bournemouth Polytechnic (1986).

Harrell L. The Effects of the Frequency and Timing of Secondary Rail Services. (M.Sc. thesis) Centre for Transport Studies, Cranfield (1987).

Harris N.G. Perceptions of Provincial Rail Services. Modern Railways, Page 14 (January 1987).

Harper W.M. Statistics. Macdonald and Evans, Plymouth (1982).

Hawkins D.I. & Tull D.S. Marketing Research - Measurement and Method. Macmillan, New York (1984).

Hayes W.L. Statistics. Holt Reinhart and Winston, Orlando Florida (1988).

Heeler, Okechuku & Reid. Attribute Importance Contrasting Measurements. Journal of Marketing Research Special Section - Measurement and Market Research, Pages 60-63 (February 1979).

Heels P. & Williams S.R. A Survey and Review of the Exeter-Barnstaple Railway Service. Polytechnic of Central London, London (February 1976).

Hedges M. & Whitehouse P. 150 Years of British Railways. Hamlyn Publishing Group Ltd., London (1980).

Hewgill M. An Examination of Minibus Design. (M.Sc. thesis) Centre for Transport Studies, Cranfield (September 1974).

Hillman M. & Whalley A. The Social Consequences of Rail Closures. Policy Studies Institute, London (1980).

Hudson, Payne & Iddiols. B.R. Sprinter Research. Consultants Report (August 1987).

Hughes Prof. J. Rural Transport Policy. Lecture notes - Centre for Transport Studies, Cranfield (date unknown).

International Standards Organisation. International Standard 2631/1 - Evaluation of Human Exposure to Whole-body Vibration, Part 1 General Requirements. I.S.O. (1985).

International Standards Organisation. International Standard 2631/3 - Evaluation of Human Exposure to Whole-body Vibration, Part 3: Evaluation of Exposure to Whole-body Z axis Vertical Vibration in the Frequency Range 0.1 to 0.63 Hz. (1985). I.S.O.

International Union of Railways. Colloquium on Lines with Low Traffic Density. Rail International, Page 34 (May 1986).

International Union of Railways. Technical Aspects of Low Cost Solutions for Lines Carrying Low Traffic Density, O.R.E. Colloquium (June 1986).

Jarman D. ScotRail Comes to Bathgate. Modern Railways, Pages 324-325 (June 1986).

Jenkins S.C & Quayle H.I. Branch Lines into the Eighties. David Charles, Newton Abbott (1980).

Joy Stewart. The Train that Ran Away. Ian Allan, Shepperton (1973).

Keen P.A. Rural Rail Services. In: Rural Transport and Country Planning, etd. Cresswell R. Leonard Hill, Glasgow (1978).

Kendall & Smith. Retranslation of Expectations: An Approach to the Construction of Rating Scales. Journal of Applied Psychology, Pages 149-155, No. 2 (1963).

Kilvington R. Railways in Rural Areas. In: International Railway Economics Studies in Management and Efficiency. etd. Button K.J. and Pitfield D.E. (1983).

Koffman J.L. A Scientific Way to Assess Passenger Comfort. Modern Railways, Pages 511- (October 1968).

Kroes & Sheldon. Article on Stated Preference Techniques. Journal of Transport Economics and Policy (January 1988).

Leach C. Introduction to Statistics, A Nonparametric Approach for the Social Sciences. Wiley, Chichester (1979).

Leatherwood J.D. et al. Ride Quality Criteria of Multifactor Environments. Source Unknown, Pages 24-31 (1979).

Laidler D. Introduction to Microeconomics. Philip Allan Pub. Ltd., Oxford (1981).

Lee M. & Round Dr. D.J. Notes of Meeting with Representatives of British Rail, Railway Technical Centre and Network South East (1988).

Leyland Vehicles, Transport & Road Research Laboratory. Passenger Problems on Moving Buses. British Leyland, Preston (September 1978).

Louviere J.J. Conjoint Analysis Modelling of Stated Preferences, A Review of Theory, Methods, Recent Developments and External Validity. Journal of Transport Economics and Policy, Pages 93-120 (January 1988).

Louviere J.J. Analyzing Decision Making, Sage University Paper Number 67, Beverley Hills California (1988).

MacBriar Dr. I.D. Bristol-Weymouth, National Bus Company Internal Report (1983).

MacBriar Dr. I.D. Cost Estimates for B.R. Passenger Services in 1982, National Bus Company Internal Report (July 1983).

Marasculio L.A. & McSweeney M. Monterey — NonParametric and Distribution-Free Methods for the Social Sciences. Brooks/Cole Publishing Co., California (1977).

Marshall J. — Guinness Factbook Rail. Guinness Superlatives Ltd., London (1985).

Meister D. (1985). — Behavioural Analysis and Measurement Methods. Wiley, New York

Miller S. — Experimental Design and Statistics. New Essential Psychology, Methuen, London (1985).

M.I.L. Research. Factors. — London and South East Rail Services: Price Elasticity of Quality Consultants Report (1982).

M.I.L. Research. — InterCity Service Quality. Consultants Report (April 1984).

Modern Railways. — The Provincial Sector Today. Modern Railways, Pages 307-310 (June 1986).

Monopolies and Mergers Board. — Report into the Provincial Sector of British Rail. H.M.S.O., London (1989).

Moseley M.J. — Accessibility: The Rural Challenge. Methuen, London (1979).

Moser & Kalton. London (1971). — Survey Methods in Social Investigation. Heinemann Educational,

M.V.A. — InterCity Rolling Stock Research - Proposal. Consultants Report (April 1985).

M.V.A. — Passenger Reaction to New Suburban Carriages. Consultants Report (May 1985).

M.V.A. — Evaluation of InterCity Rolling Stock Improvements. Consultants Report (May 1986).

M.V.A. — Interim Report, Network South East Quality of Service Study. Consultants Report (January 1988).

Nash C.A. — Economics of Public Transport. Longman Group Ltd., Harlow (1982).

Nash C.A. — Prime User Costing. In: Transport UK 1985. Pages 113-119, Policy Journals, Newbury (1985).

Nash C.A. & Preston J. — Some Guidelines for Evaluating New Local Rail Stations. Institute for Transport Studies, University of Leeds (June 1986).

Nock O.S. — Branch Lines. B.T. Batsford Ltd., London (1957).

Norusis M.J. — SPSS-X Advanced Statistics Guide. McGraw-Hill, New York (1988).

Open University.	Modelling for Understanding: A Case Study in Econometrics. Open University Press, Milton Keynes (1976).
Open University.	Data Collection Procedures, Research Methods in Education and the Social Sciences. Open University Press, Milton Keynes (1979).
Oppenheim A.N.	Questionnaire Design and Attitude Measurement. Heinemann Educational Books Ltd., London (1966).
Pearmain D.	The Measurement of Passenger Valuations towards Rail Station Facilities. Oxford Polytechnic Report, Oxford (January 1989).
P.E.I.D.A.	The Settle and Carlisle Railway. P.E.I.D.A., Edinburgh (1985).
Perren B. 1986).	New D.M.U.s Enter Service, Modern Railways, Pages 321-323 (June
Policy Studies Institute.	The Future of Rural Railways. Conference Proceedings P.S.I., London (1981).
Pollard Dr. M.G.	Passenger Tolerance of High Speed Curving. Railway Gazette International, Pages 870-873 (November 1984).
Preston J.	The Evaluation of New Local Rail Stations in West Yorkshire, (Ph.D. thesis) Institute for Transport Studies, University of Leeds (1987).
Prideaux Dr. J.	InterCity - Passenger Railway Without Subsidy. Lecture given to the Railway Study Association (24th May 1989).
Planning and Transport and Research and Computation (International) Co.	The National Rail Network in Britain. P.T.R.C. Summer Annual Meeting, University of Sussex, Brighton (1983).
Rail Power.	Wide-body Railbus Enters Service. Rail Power No. 47, Page 17 (October 1985).
Rail Power.	Metro-Cammell's Prototype D.M.U. Rail Power No.47, Pages 14-15 (October 1985).
Rail Power.	British Rail Expands Railbus Order. Rail Power No. 48, Page 7 (July 1986).
Railway Development Society.	Bring Back the Trains. R.D.S., London (1984).
Railway Development Society.	A New Future for Local Trains. Promotional Leaflet, R.D.S., London (date unknown).
Railway Gazette	Four Elements in the Low Cost

International 1981).	Railway. Railway Gazette International, Pages 130-131 (February
Railway Gazette International.	Diesel-Hydraulic Trains part of Rural Economy Package. Railway Gazette International, Pages (November 1981).
Railway Gazette International.	B.R. Tries Another Railbus Tack. Railway Gazette International, Page 561 (July 1982).
Railway Gazette International.	Radio Data Links can replace Low-Intensity Signalling. Railway Gazette International, Pages 909-912 (November 1982).
Railway Gazette International.	Railbus is Last Chance to get Local Services Right. Railway Gazette International, Page 53 (January 1983).
Railway Gazette International.	D.M.U. Prototypes on Trial. Railway Gazette International, Page 442 (July 1984).
Railway Gazette International	Highland Rail gets Results. Railway Gazette International, Page 360 (May 1985).
Railway Invigoration Society.	Can Bus Replace Train. R.I.S., London (1977).
Reeves R. (etd.)	Product Development Cases Collection (Advanced Passenger Train), School of Mechanical Engineering, Cranfield (date unknown).
Reid Sir R.B. (June 1982).	Railway Business Management by Sector. Rail International, Page 3
Research Projects.	Replacement of pre-war E.M.U. stock. Consultants Report (December 1968).
Richards K. Economics	Economics of the Cambrian Coast Line. Journal of Transport and Policy, Page 308 (1972).
Round Dr. D.J.	The Current Work of the Vehicle/Track Systems Group (Track Research Unit). B.R. Internal Report, Railway Technical Centre, Derby (November 1987).
Serpell Committee.	Railway Finances. H.M.S.O., London (1983).
Social and Community Planning Research.	Depth Interviews and Group Discussions. S.C.P.R., London (1972).
St John Thomas D.	The Rural Transport Problem. Routledge and Kegan Paul, London (1963).

St John Thomas D. The Country Railway. Book Club Associates, London (1977).

St John Thomas D, & Whitehouse P. The Great Days of the Country Railway. David and Charles, Newton Abbott (1986).

St John Thomas D. A Regional History of the Railways of Great Britain, Volume One - The West Country. David and Charles, Newton Abbott (date unknown).

Steer Davies & Gleave et al. B.R. Passenger Rolling Stock LEV/140/210: Stage 2 Report. Consultants Report (November 1981).

Steer Davies & Gleave. Passenger Reaction to the Class 140 Unit on Central Wales Line. Consultants Report (August 1982).

Steer Davies & Gleave. B.R. Passenger Rolling Stock LEV/140/210: Stage 3 Report. Consultants Report (January 1983).

Steer Davies & Gleave. London-Bournemouth Rolling Stock Evaluation. Consultants Report (April 1983).

Steer Davies & Gleave. Passenger Reactions to a Mark One/Leyland Body Passenger Coach. Consultants Report (March 1984).

Stewart M.L, et al. Ride Quality Ratings as a Function of Sensory Input, Ergonomics No 4, Pages 399-405 (1979).

Taylor W. (etd.) The Vibration Syndrome. Academic Press, London (1974).

T.D.M. MacMeter Mark II - Operating and Service Manual. T.D.M., Long Eaton (January 1986).

T.D.M. MacMeter Mark II - Passenger Ride Comfort Meter. T.D.M., Long Eaton (date unknown).

Thomas R.L. Introductory Econometrics, Theory and Applications. Longman Group Ltd., London (1985).

Towriss Dr. J.G. The New Car Buyer: The Determinants of Choice, Centre for Transport Studies, Cranfield (July 1981).

Towriss Dr. J.G. A New Approach to the Use of Expectancy Value Models, Journal of the Market Research Society, Pages 63-75 (January 1984).

Transmark. Highland Rail Study. Consultants Report (1984).

Transport & Road Research Laboratory. The Demand for Public Transport. T.R.R.L., Crowthorne (1980).

Transport Users Consultative Committee. Annual Report. H.M.S.O., London (September 1987).

Tuck M. How do we Choose - A Study in Consumer Behaviour. Methuen, London (1976).

Van Wouden E.F.H. Simplified Operation of Secondary Lines, Rail International, Pages 5-9 (January 1987).

Vorhees A.M. & Associates. Proposed Investment in Edinburgh (Waverley) Station. A.M.V., London (March 1971).

Winters D.I, An Examination of Bus Design. (M.Sc. thesis) Centre for Transport Studies, Cranfield (1979).

Appendices

Glossary of Terms.

Cant	The amount a railway track is banked, to reduce the effects of lateral acceleration on passengers.
Cant Deficiency	The amount of extra cant needed on a corner to offset lateral forces at a given speed.
Conjoint Analysis	Examines utility and preference formulation to understand choice decisions.
Contamination	A rating is affected by otherobjects that are in some way related to it - the inability to isolate feelings for one object.
Engineering Ride	Ride quality measured using I.S.O. weighted MS-2 R.M.S.
Halo	The rating of an object is biased by an overall un/favourableness towards the object.
Hyper-Sensitivity	It is suggested that respondents who know they will be assessing an experience, before it has happened, will produce more critical ratings.
Interference	The overall bias in a rating, some combination of Halo, Newness and Contamination effects.
Newness	An object's rating is biased by a feeling of novelty, extreme cleanliness etc.
Priority Evaluator	Establishes the relative importance of something.

Engineering Levels Of Ride For Whole Route.

Lateral and vertical I.S.O weighted R.M.S. ride measurements in guards van over coupled bogey in class 101 (powered) driving coach. 24/4/89.

Bristol Temple Meads 09.00 hours to Weymouth (southbound).

SECTION	LATERAL	VERTICAL	LOAD	SPEED
Bristol Temple Meads - Keynsham	17	37	Full	
Keynsham - Oldfield Park	14	33	"	
Oldfield Park - Bath Spa	5	18	"	Slowed
Bath Spa - Bradford On Avon	15	40	"	Caution
Bradford On Avon - Trowbridge	11	36	"	
Trowbridge - Westbury	9	34	"	
Westbury - Frome	14	37	"	
Frome - Bruton	16	44	"	
Bruton - Castle Cary	11	32	"	
Castle Cary - Yeovil Pen Mill (P.M)	22	38	"	Slowed
Yeovil P.M. - Maiden Newton	15	37	"	
Maiden Newton - Dorchester West	19	41	"	
Dorchester West - Upwey	12	37	"	
Upwey - Weymouth	10	32	"	

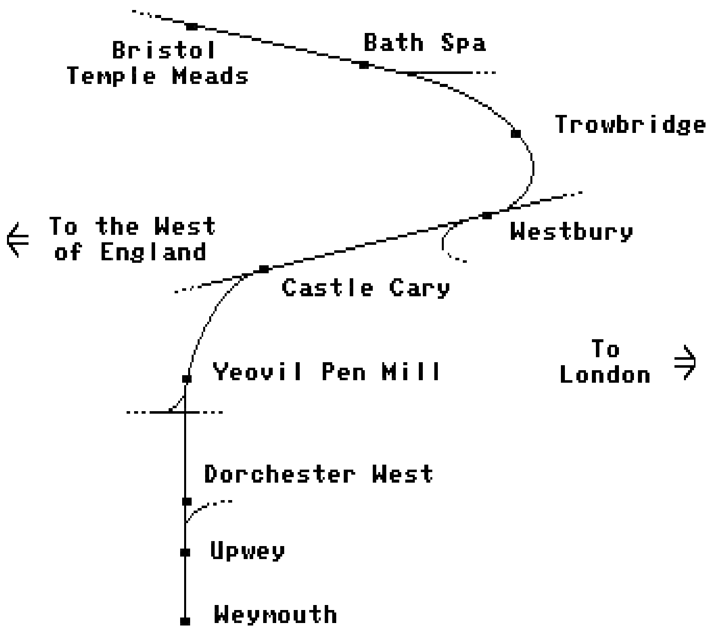
Bristol Temple Meads 16.08 hours to Weymouth (southbound).

SECTION	LATERAL	VERTICAL	LOAD	SPEED
Bristol Temple Meads - Keynsham	14	40	Empty	
Keynsham - Oldfield Park	11	42	"	
Oldfield Park - Bath Spa	10	33	"	
Bath Spa - Bradford On Avon	12	39	Full	
Bradford On Avon - Trowbridge	9	34	"	
Trowbridge - Westbury	9	31	"	
Westbury - Frome	14	36	"	
Frome - Bruton	14	39	"	
Bruton - Castle Cary	10	32	"	
Castle Cary - Yeovil P.M.	23	40	"	
Yeovil P.M. - Maiden Newton	14	37	1/2 Full	
Maiden Newton - Dorchester West	18	42	"	
Dorchester West - Upwey	12	42	"	
Upwey - Weymouth	12	36	"	

The Survey Area

The bulk of the surveys were conducted on sections of the Bristol-Weymouth line. This line is fairly typical of Provincial's longer secondary routes. Under the recommendations of the Beeching report it was not scheduled for closure, though many of the stations along it were. It was soon realised that these proposals were not practical and a number of stations were reprieved.

MAP OF SURVEY AREA.



The line is currently managed by British Rail's Provincial Sector (Western) and divided between the Western and Southern Regions at Dorchester. The only traffic that remains is a local passenger service and a few summer excursion trains. Traffic is seasonal and trains can become very crowded in summer. North of Trowbridge passengers flows are peaky and dominated by short flows to/from Bristol and Bath. South of Trowbridge longer distance trips are more prominent.

The line has a two-hourly service calling at most stations, though some of the intermediate stations are request stops. The trains used on the line were usually Diesel Multiple Units (D.M.U's) of the, "Modernisation Plan" (1955-65) era. These were replaced, during the study, by class 150 and 155 Sprinters. There is some competition for the line, both from National Express and local bus/coach operators - some of whom are subsidised by Dorset County Council.

The line has considerable potential. It is expected that unless there are cuts in British Rail's grant, or other unforeseen difficulties the line's future will be secure.