
Published paper

Blana, E. (1996) *Driving Simulator Validation Studies: A Literature Review.* Institute of Transport Studies, University of Leeds, Working Paper 480

UNIVERSITY OF LEEDS
Institute for Transport Studies

ITS Working Paper 480 *ISSN 0142-8942*

December 1996

Driving simulator validation studies: A literature review

Evi Blana

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This literature review is part of the "Driver performance in the EPSRC driving simulator: a validation study" project funded by EPSRC. It focuses mainly on driving simulator validation studies with regard to driver behaviour. Various approaches, methodologies and criteria have been proposed until today regarding the behavioural and physical validation of a driving simulator. At the same time, a number of behavioural validation studies have been conducted, with or without taking into account the proposed validation approaches. The author considered necessary this literature review because according to her knowledge there was no other published review which examined thoroughly the link between theory (proposed validation approaches and methodologies) and practice (validation studies on driving simulators). Most of the recent behavioural validation studies have been focused on the absolute and relative validity of the simulator without taking into consideration the issue of the face validity.

The format of this paper will be as follows. A small introduction to driving simulators and their usefulness will be presented first followed by the necessity of validating them. The existing validation approaches, methodologies and criteria will be analysed and earlier and recent behaviour validation studies will be reviewed and compared in detail. These studies will be classified according to the driver behaviour levels and driving performance (as they will be defined) and then will be assessed according to the validation criteria. Emphasis will be given to the interpretation of the findings from these comparisons, and in particular to their applicability in real road traffic situations.

2. THE NEED FOR DRIVING SIMULATORS

Driving simulators were first developed for the training of a large number of personnel in the tactical use of war machinery during the Second World War. In the early 1960's, they were applied in the research field to study driver behaviour and his/her interaction with the vehicle and the road environment (Roberts, 1980). Due to rapid progress of the state of the art in visual displays and computer technology by 1975 at least sixteen driving simulators were operating throughout the United States using different techniques for the generation of the visual field and two in Europe (Allen, Klein and Ziedman, 1979). The latest trend to the development of driving simulators (after 1985) is to fulfil the specific needs of a particular group, whether this is an automotive industry, a private research institute or a university.

The main application areas of today driving simulators have been to investigate acceptability issues of innovative transport elements (e.g. road design, in-vehicle device), to evaluate the safety concept (e.g. possible increase of accidents due to new road design, in-vehicle device) as well as the credibility and transferability of the simulator results to the real world. Driving simulators have been used as research aids in a number of civil engineering, transport, psychology and ergonomics fields such as: innovative road design (e.g. testing the design of new tunnels, innovative highway design and road delineation, traffic calming); intelligent transport systems (e.g. new in-vehicle navigation systems, Head-Up-Displays, active pedals); impaired driver behaviour (driving behaviour affected by drugs, alcohol, severe brain damage, fatigue) and vehicle dynamics and layout (e.g. testing ABS, 4-wheel drive; vehicle interior design).

The main advantage of driving simulators is that they can provide an inherently safe environment for driving research, which can be easily and economically configured to

investigate a variety of human factors research problems. They make it possible to control experimental conditions over a wider range than field tests and can be easily changed from one condition to another. They are linked to digital computer systems which can further provide on-line data processing, formatting and storage and the reduction and compact arrangement of data.

On the other hand, driving simulators provide drivers with an artificial environment which could never be the same as the real one. For example, even in the most advanced driving simulators the longitudinal and lateral accelerations are limited (e.g. VTI driving simulator: Nilsson, 1989; Daimler-Benz driving simulator: Drosdol and Panik, 1985) and only parts of the extremely complicated transport system can be simulated until today. The differences between the simulated and the real driving environment may influence subjects' driving behaviour and performance, hence any performance measures observed in a driving simulator may differ from the same measures observed during real driving. Therefore, the issue of evaluating the driving simulators emerges in order for them to produce transferable, reliable, and valid results between the two environments.

3. EVALUATION OF DRIVING SIMULATORS

The evaluation of driving simulators could be separated into three stages: a) the transferability of the results obtained from a driving simulator to real world; b) the reliability of a driving simulator and c) the validity of a driving simulator. The first stage is crucial and rather necessary for the training simulators (either driving or flight simulators) and it has been extensively investigated for flight simulators. The reason why the second stage has not been given a lot of attention from the researchers and is mentioned very rarely is because a valid driving simulator is always reliable too, where the vice versa does not apply. The third and most important stage for any simulator, is the issue of validity and it is examined here thoroughly.

3.1. TRANSFERABILITY AND RELIABILITY OF RESULTS OF DRIVING SIMULATORS

The issues of transferability and reliability of results obtained from a driving simulator will not be examined in detail since there are not the main topic of this paper, however definitions of these terms and examples relative to driving simulators will be given in the following paragraphs.

3.1.1. The issue of transferability

The issue of transferability of results from the simulated to the real environment has always been of critical importance for the training rather than the research simulators. Especially for the flight simulators, it was investigated since their first development and usage as pilot training devices in the aircraft and military industry. Caro (1973) in his investigation of different training techniques, he concluded that the potential training value of a flight simulator depends probably more on a proper training program than on the actual realism of the simulator. He also claimed that "transfer of training from a device to an aircraft is limited to the tasks which can be performed in the device. But, whether that limit is reached is a function of the way in which the device is used". Valverde (1973) in his review of flight simulator transfer of training studies concluded that in order to eliminate contradictory results from these studies, one must take into serious consideration the criterion measures; the

individual differences between subjects as well as their motivation and attitudes towards the simulator; the attitude, ability and motivation of the instructor; and the instructional sequence.

For a driving simulator in order to be valid it should allow at least the transfer of basic driving skills from a real driving environment to a simulated and at the same time it should present the subject with realistic visual and auditory cues and traffic scenarios. Since the objective of this paper is the behavioural validity of research driving simulators, transferability will not be examined further.

3.1.2. The issue of reliability

In terms of psychology reliability is the consistency with which a test measures what it measures (Gleitman, 1991). In determining the reliability of a measuring instrument (e.g. a driving simulator), it is assumed that the instrument is measuring a relatively stable characteristic.

Because a driving simulator is a very complicated system, an aggregation of a number of subsystems, it is clear that we cannot claim overall reliability of a simulator but reliability of each of its subsystems. For example, the vehicle dynamics model of a driving simulator measures the braking force via the braking system of the simulated vehicle. If the braking system of the simulated vehicle is half working then we will receive consistent results, the braking force, but not the correct ones. The braking system will be reliable but not valid. This means that high reliability alone does not guarantee that a test (here the braking system of the simulator) is a good measuring device. But, at the same time, a driving simulator is also a man-in-the-loop device which means that one has to check both the physical as the behavioural reliability of the simulator. The author believes that although the physical reliability of the simulator and its subsystems can be measured and tested easily, the behavioural reliability is not only difficult to measure or test it, but even more to identify it. Unreliability can be a result of measurement errors produced by temporary internal (e.g. for the behavioural reliability: low motivation, temporary indisposition of the subjects) or external (e.g. for the behavioural reliability: a distracting or uncomfortable testing environment) conditions (Aiken, 1971). More critical to a test's reliability is its validity, and since it is one of the objectives of this study, it will be thoroughly examined in the next paragraphs.

3.2. VALIDITY OF DRIVING SIMULATORS

Defining the validity of a driving simulator is a multi-disciplinary and complicated task. Mudd (1968) defined validity as the way in which the simulator "reproduces a behavioural environment" where according to Allen et al (1991) "Validity is only defined to a specific research question". Rolfe et al (1970) stated that "The value of a simulator depends on its ability to elicit from the operator the same sort of response that he would make in the real situation". According to Leonard and Wierwille (1975) "simulator validation is a problem of obtaining parallel measures in full-scale and in simulation and bringing these two sets of measures into correspondence". It is clear that the term "validity of a driving simulator" is not precisely defined and it needs further specification. On the other hand, validity from the standpoint of psychology is widely used for the assessment of psychological tests, it is clearly defined and there are already standards relative to the validity of a test (e.g. APA, 1985). However, there is a major problem in this instance. The driving simulators is a man-in-the-loop system and human performance is confounded with system performance. This problem has been recognised from the early sixties. Ebel (1961) proved that very few psychological tests used in clinics and industry had satisfactory validity data when used in the simulators,

which implied that the current concepts of validity may not be adequate. Still, the author judged as substantive to mention validity in terms of psychology and attempt to apply these terms to the driving simulator validity and confirm or not the above findings.

The validity of a test is defined as the extend to which it measures what it is supposed to measure (Gleitman, 1991). Unlike reliability which can only be influenced by unsystematic errors of measurement, validity can be affected by both unsystematic and systematic (constant) errors, i.e. a test may be reliable without being valid, but it cannot be valid without being reliable. This means that reliability is a necessary but not a sufficient condition for validity. Primarily, all procedures for determining test validity are concerned with the relationships between performance on the test and other independently observable facts about the behaviour characteristics under consideration. (Anastasi, 1988)

Validity has been classified into different categories. Tiffin and McCormick (1965) classified validity into four categories: a) concurrent, b) predictive, c) content and d) construct. The first two categories are tested using statistical or quantitative methods because they rely on numerical performance data for analysis whereas the last two by using logical or qualitative methods because they use judgement type data for analysis. Gleitman (1991) classified validity as previous but only using the terms of predictive and construct validity. The American Psychological Association (APA) (1985), in its proposal for the technical standards for test construction and evaluation, grouped validity evidence into three categories: a) the construct related, b) the content related and c) the criterion-related evidence of validity, which includes the predictive and concurrent validity. The same grouping is used by Anastasi (1988). According to APA (1989) standards, an ideal validation should include several types of evidence, comprising of all three traditional categories. The quality of the evidence is of primary importance and a single line of solid evidence is preferable to numerous lines of evidence of questionable quality. Gathering evidence may sometimes involve examining not only the present instrument in the present situation, but also the available evidence on the use of the same or similar instruments in similar situations. This working paper will refer to validity classification according to the APA and will try to associate this classification of validity with the validity of a driving simulator.

3.2.1. Criterion-related validity

According to McCoy (1963) the first and most important consideration in setting up an experimental investigation is the development of precise criteria. This is fundamental to selecting the parameters for measurement, and the measures to use, as well as providing the frame of reference in which validation will be attempted. The criterion-related validity indicate the effectiveness of a test in predicting an individual's performance in specified activities. For this purpose, performance on the test is checked against a criterion, that is, a direct and independent measure of that which the test is designed to predict (Anastasi, 1988). The choice of the criterion and the measurement procedures used to obtain criterion scores are of central importance. The value of a criterion-related study depends on the relevance of the criterion measure that is used. Whether a given degree of accuracy is judged to be high or low or useful or not useful depends on the context in which the decision is to be made (APA, 1985).

When establishing the criterion-related validity of a driving simulator, we need to consider the existing driving simulator validity criteria and if they are not adequate, we have to develop new more precise criteria but more over to define precisely the experiment, which in a way plays the role of the test. But, before defining any criteria or test protocols, we should take

into account the fact that there is both physical and behavioural validity of a simulator. The physical validity is absolute necessary condition for the behavioural validity. Behavioural validity can be checked if and only if the physical validity has already been tested and verified. Physical validity can not only enhance the behavioural validity but also minimise the internal variables that may affect behavioural validity.

Criterion-related validity is often used in local validation studies, in which the effectiveness of a test for a specific program is to be assessed. According to Anastasi (1988) the application of criterion-related validation is not technically feasible for small samples (40-50 cases). Aiken (1971) concluded that the correlation between the test and an external criterion measure can never be greater than the square root of the parallel forms reliability coefficient of the test. Factors affecting criterion-related validity can be group differences, test length, criterion contamination, base rate and incremental validity (Aiken, 1971).

When applied to driving simulators, it can also be affected by group differences, i.e. if we only use male instead of both male and female subjects for the simulated experiments (there is evidence of different driving patterns between males and females on real road); if we only use young instead of both young and old subjects (there is evidence of different driving patterns between young and old drivers on real road and lately in the simulators too), therefore it would be preferable if the sample size (n) is weighted for gender and age and could be greater than fifty (n=50) to avoid subjects' variability. It can also be affected by test length, i.e. the experiment in the simulator should not be so long in order to avoid fatigue, monotony, and simulator sickness.

The criterion-related validity can be distinguished to predictive and concurrent validity. A predictive study obtains information about the accuracy with which early test data can be used to estimate criterion scores that will be obtained in the future. A concurrent study serves the same purpose, but it obtains prediction and criterion information simultaneously. A decision theory framework can be used to judge the value or utility of a predictor test. One judgement can be that the most important error to avoid is a false positive -selecting someone who will subsequently fail. Another judgement focuses on avoiding a false negative -not selecting people who would have succeeded. The relative cost assigned to each kind of error is again a value judgement; depending on that judgement, the subsequent interpretation of the utility of testing may differ (APA, 1985). The logical distinction between predictive and concurrent validation is based, not on time, but on the objectives of testing (Anastasi, 1988).

Predictive validity can be achieved by evaluating the effectiveness of the simulator in predicting certain driver performance in the future from present performance data is tested requiring a follow-up study (Leonard and Wierwille, 1975). The simulator is used as the test or measure to predict future driver behaviour. This can be measured by the correlation coefficients. The appropriate criterion should be the genuine road used driving behaviour. Maybe it would be more practical to define a local or situation specific criterion related evidence of driving simulator validity. For example relative to driving speed, we should define:
a) differences in measuring the speed on the road and in the simulator;
b) the type of simulator used for the experiment;
c) the type of subjects used;
d) the time the real road and the simulated experiment took place.

For driving simulators, concurrent validity can be achieved by comparing real road and simulated data (Leonard and Wierwille, 1975). Generalisation of a driving simulator validity

is limited because validity is usually applicable for a particular task and a particular driving simulator.

3.2.1.1.Predictive validity

Predictive validity is evaluated by showing how will predictions made from a test or measure are confirmed by subsequent observation. Commonly, it is expressed in terms of a simple correlation between test scores or measures and some criterion scores or measures. An obvious problem here is the determination of precise criteria (McCoy, 1963).

One index of a test's validity is the success with which it makes such prediction. This is usually measured by the correlation between the test score and some appropriate criterion. The correlation coefficient is a mathematical expression that summarises both the direction and the strength of the relationship between the two measures. It varies between ±1.00 and it is symbolised by r. The ± sign indicates the direction of the relationship where the strength of the correlation is expressed by its absolute value. When r=1, i.e. when the correlation is perfect and prediction is error-free, then there is no more variation at all around the line of best fit. But, the fact that two variables are correlated says nothing about the underlying causal relationship between them. Validity coefficients are just the correlations between test scores and criteria and their magnitude depends upon the range of ability within the group in which they are determined. As this range is narrowed, the correlation between test score and criterion declines. This relationship holds for all correlation coefficients (Gleitman, 1991).

The interpretation of a validity coefficient must take into account a number of concomitant circumstance, therefore the obtained correlation should be high enough to be statistically significant at some acceptable level, such as the 0.01 or 0.05 levels. Before drawing any conclusions about the validity of a test we should be reasonably certain that the obtained validity coefficient could not have arisen through chance fluctuations of sampling from a true correlation of zero. When a significant correlation between test scores and criterion has been established, we need to evaluate the size of the correlation on the light of the uses to be made of the test (Anastasi, 1988).

3.2.1.2.Concurrent validity

The only difference between concurrent and predictive validity is the point in time when the validating criteria are determined. The measure to be validated and the criterion measure are usually taken simultaneously or at about the same time. Generally the reason for seeking concurrent validity is to substitute one measure for another. According to McCoy (1963) concurrent validity can be evaluated by showing how well test scores or measures correlate with concurrent status or performance.

3.2.1.3.Validity generalisation

Earlier the judgements about generalisation were based upon non-quantitative reviews of the literature. Later, quantitative meta-analytic techniques were used. The results of validity generalisation studies can be used either to draw scientific conclusions and/or to use the results of validity evidence obtained from prior studies to support the use of a test in a new situation. The latter use raises questions about the degree to which validates are transportable to a specific new situation. If generalisation is limited, then local criterion-related evidence of validity may be necessary in most situations in which a test is used. If generalisation is extensive, then situation-specific evidence of validity may not be required. In conducting

studies of the generalisability of validity evidence, the prior studies that are included may vary according to several situation facets. Some of the major facets are differences in the way the predictor construct is measured; the type of job or curriculum involved; the type of criterion measure; the type of test takers and the time period in which the study was conducted. A major objective of the study should be to decide whether variation in these facets affects the generalisability of validity evidence (APA, 1985).

3.2.1.4.Validity Standardisation

To evaluate a test, we need one further item of information in addition to its reliability and validity. We have to know something about the group on which the test was standardised. A crucial requirement in using tests is the comparability between the subjects who are tested and the standardisation sample that yields the norm. If these two are drawn from different populations, the test scores may not be interpretable. The standardisation of any psychological or educational assessment instrument requires administering the instrument to a large sample of individuals (the standardisation sample) selects as representative of the target population of persons for whom the instrument is intended (Aiken 1971).

We are still too far from a driving simulator behavioural validity standardisation for the following reasons
a) there are no standard tests to assess driver behaviour either on the road or in the simulator;
b) although there are some standard tests used for the automotive industry for checking the capabilities of new cars, they are rarely applied for studies on the simulator;
c) usually for each behavioural validation study, there are major differences on the type of simulators, subjects, test protocol, data collection methods.

We need first to create the standard tests or improve/enhance the existing ones relative to driver performance, then apply them both on real road and in the simulator using large number of subjects on both environments (in order to avoid in-between subjects variability) and finally conclude about the validity standardisation.

3.2.2. Content-related validity

Content-related validity demonstrates the degree to which the sample of items, tasks, or questions on a test are representative of some defined universe or domain of content. The domain under consideration should be fully described in advance, rather than after the test has been prepared (Anastasi, 1988). The methods often rely on systematic observation of behaviour combined with the expert judgements but also certain logical and empirical procedures can be used to assess the relationship between parts of the test and the defined universe. Sometimes algorithms or rules can be constructed to generate items that differ systematically on various domain facets, thus assuring representiveness. Also. The first task of test developers is to specify adequately the universe of content that a test is intended to represent, given the proposed uses of the test (McCoy, 1963; APA, 1985).

Content validity depends on the relevance of the individual's test responses to the behaviour area under consideration, rather than on the apparent relevance of item content. It is also important to guard against any tendency to overgeneralize regarding the domain sampled by the test. Another difficulty arises from the possible inclusion of irrelevant factors in the test scores. The end product of the test-development procedures consists of the items actually included in the final version of the test. The manual should provide information on the content

areas and the skills or instructional objectives covered by the test, together with some indication of the number of items in each category.

When test users consider using an available test for a purpose other than that for which the test was developed originally, they need to judge the appropriateness of the original domain definition for the proposed new use. Another important task is to determine the degree to which the format and response properties of the sample of items or tasks in a test are representative of the universe. Methods classes in this category should often be concerned with the psychological construct underlying the test as well as with the character of test content. There is often no sharp distinction between test content and test construct. Content-related evidence of validity is a central concern during test development, thus inferences about content are linked to test construction as well as to establishing evidence of validity after a test has been developed and chosen for use. (Anastasi, 1988).

The critical question for content-related validity is if the driver performance measures that we usually choose as representable of driving performance in real life are also representable of driving performance in a simulator. But before establishing the behavioural content-validity of a simulator, the physical content-validity has to be established, i.e. if the simulator as an aggregation of different subsystems is representative of a real car on a real road. Leonard and Wierwille (1975) suggested that content validity can be achieved by using the subjective opinion of how well the simulator resembles a real road automobile, which according to the author is face validity.

3.2.2.1.Face validity

Content validity should not be confused with face validity. The latter is not validity in the technical sense; it refers, not to what the test actually measures, but to what it appears superficially to measure. Face validity pertains to whether the test "looks valid" to the examinees who take it, the administrative personnel who decide on its use and other technically untrained observers. When referring to driving simulators, it pertains to whether the simulator looks valid to the subjects driving it, to the people deciding on its use and on further investment, to the researchers who use it for their experiments.

A low face validity does not necessarily directly affect the validity of results. Yet, if it affects, e.g. subject's motivation, this in turn might affect validity. The general result of studies that compared different types and amounts of cues in driving simulators is that most cues increase face validity. A driving simulator is more realistic with more complex visual information, whereas the effect of a moving-base increases with the number of degrees of freedom. Yet, these studies do not relate face validity to absolute or relative validity of simulators to address different types of research questions. The functional validity questions the validity of the results regarding the resemblance between the behaviour in the simulator and the real task environment. Face validity should never be regarded as a substitute for objectively determined validity. As Harms et al (1996) concluded, increasing the face validity of the VTI driving simulator, it didn't necessarily enhance the overall behavioural validity of the simulator.

Although this term of validity may cause some confusion, it is a desirable feature of tests. Face validity can often be improved by merely reformulating test items in terms that appear relevant and plausible in the particular setting in which they will be used. It cannot be assumed that improving the face validity of a test will improve its objective validity. Nor can it be assumed that when a test is modified so as to increase its face validity, its objective validity remains unaltered. (Anastasi, 1988).

3.2.3. Construct-related validity

The construct validity is the extent to which the performance on the test fits into a theoretical characteristic–or construct–about the attribute the test tries to measure (Aiken, 1971). This characteristic is called "construct" because it is a theoretical construction about the nature of human behaviour (APA, 1985). Each construct is developed to explain and organise observed response consistencies. It derives from established interrelationships among behavioural measures and it requires the gradual accumulation of information from a variety of sources. Any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestation represent appropriate evidence for this validation. (Anastasi, 1988).

Establishing the construct validity of a measure is a problem distinct from that of using that measure in predicting a second measure, although the latter can often contribute to construct validation. Evidence from content- and criterion-relation validation studies contributes to construct interpretations. The choice of which of one or more approaches to use to gather evidence for interpreting constructs will depend on the particular validation problem and the extent to which validation is focused on construct meaning (APA, 1985). Messick (1980) argued that the term validity, insofar as it designates the interpretative meaningfulness of a test, should be reserved for construct validity and other procedures with which the term validity has been traditionally associated should be designated by more specifically descriptive labels. Thus, content validity can be labelled content relevance and content coverage, to refer to domain specifications and domain representiveness, respectively. Criterion-related validity can be labelled predictive utility and diagnostic utility, to correspond to predictive and concurrent validation.

Construct validity depends on:
1) Experts' judgements that the content of the test pertains to the construct of interest;
2) An analysis of the internal consistency of the test;
3) Studies of the relationships, in both experimentally contrived and naturally occurring groups, of test scores to other variables on which the group differ;
4) Correlations of the test with other tests and variables with which the test is expected to have a certain relationship and factor analyses of these intercorrelations;
5) Questioning examinees or raters in detail about their responses to a test or rating scale in order to reveal the specific mental processes that occurred in deciding to make those responses (Aiken, 1971).

For driving simulators, construct validity can be achieved when we test if the simulator's data can be examined relative to identical hypothetical constructs used in other driving research (Breda et al, 1972; Leonard and Wierwille, 1975).

3.2.3.1.Convergent and discriminant validation

Construct validity, can be obtained if the test has high correlations with other measures (or methods of measuring) of the same construct (convergent validity) and low correlations with measures of different constructs (discriminant validity). Evidence regarding to the convergent and discriminant validity of an instrument can be possessed by comparing correlations between measures of the same construct using the same method; different constructs using the same method; the same construct using different methods; and different constructs using different methods.

Factor analysis is particularly relevant to construct validity because it is primarily used for analysing the interrelationships of behaviour data. A major purpose of factor analysis is to simplify the description of behaviour by reducing the number of categories from an initial multiplicity of test variables to a few common factors or traits. After the factors have been identified, they can be utilised in describing the factorial composition of a test. Each test can thus be characterised in terms of the major factors determining its score, together with the width or loading of each factor and the correlation of the test with each factor. Such a correlation is known as the factorial validity of the test. Correlations between a new test and similar earlier tests are sometimes cited as evidence that the new test measures approximately the same general area of behaviour as other tests designated by the same name. Unlike the correlations found in criterion-related validity, these correlations should be moderately high, but not too high. If the new test correlates too highly with an already available test without such added advantages as brevity or ease of administration, then the new test represents needless duplication. Correlations with other tests could also be employed to demonstrate that the new test is relatively free from the influence of certain irrelevant factors (Anastasi, 1988).

Factor analysis is usually used in traffic engineering to determine which are the most important measures of driving performance on the real road. It could also be employed to identify the respective measures of driving performance on the simulator and then compare if these measures are the same for both environments (real road and simulator) and test if they are highly correlated.

3.2.4. CONCLUSIONS

McCoy (1963) investigated the applicability of the psychological terms of validity to man-machine systems and concluded that "the concepts of validity currently adhered are not of practical use to the human engineer interested in determining, quantitatively, the degree of validity attained in such a measurement system". Hoffman and Joubert (1966) concluded that opinion data, such as that used to determine content validity may not always be reliable indices of system performance. Leonard and Wierwille (1975) concluded that the "ultimate method for determining validity is [by] determining the degree of concurrent or predictive validity" and proposed a theory "using the basic concept of concurrent validity applied to measured human performance".

A literature review of the typical psychological measurement assessment theory and its application to driving simulators showed that it has been proven extremely difficult to apply the psychological definitions of validity to driving simulators. The author has to agree with Ebel (1961) and McCoy (1963) that these terms are not adequate for the simulators, because they are man-machine interacted systems and this interaction is too far complicated to be explained by these terms and new standards and procedures must be developed for the overall validation of a driving simulator.

4. A REVIEW OF DRIVING SIMULATORS VALIDATION APPROACHES, METHODOLOGIES AND CRITERIA

4.1. DRIVING SIMULATOR VALIDATION APPROACHES

A literature review on the existing validation approaches or methodologies revealed that the first approach to the validation of simulators was made by Mudd (1968) and McCormick (1970). They distinguished the validation as **behavioural correspondence** (between the behaviour of the subject in the simulator and on the real vehicle) and **physical**

correspondence (between the simulator and the vehicle and includes, for example, layout and dynamic characteristics). The two aspects of validity do not have to be necessarily related. Generally the behavioural correspondence is assumed to be more important for the validity of a simulator for a specific task. The earlier simulator studies mentioned the physical correspondence only and paid less attention to the behavioural correspondence. Behavioural validation studies of simulators started around 1970 and referred to driving simulators with limited graphics presentation and computing abilities (Allen and O' Hanlon, 1979).

A very similar to the above approach was later proposed by Brown (1975) and Blaauw (1982). Blaauw made the distinction between correspondence in driver's behaviour and driver's performance regarding the validity of a driving simulator. In addition to the comparison of driving behaviour in the simulator and on the road under the same conditions, he also proposed the comparison of performance differences between the simulated and the real world under similar conditions. This approach was widely used by other researchers for the validation of their driving simulator (e.g. Green and Reed, 1995: validation of the UMTRI driving simulator; Harms, 1993 and Alm, 1995: validation of the VTI driving simulator).

A similar technique to the behavioural and physical correspondence for validating the driving simulators was introduced by Bertollini et al (1986): the **closed-loop** and the **open loop** techniques. Closed-loop techniques attempt to show how performance, performance trends, and subjective ratings correspond between simulator, full-scale vehicle or other driving research data. Open-loop techniques attempt to verify that the models accurately represent the vehicle response without the driver.

The methodology proposed by Allen et al (1991) is also similar to the methodologies described above. Their contribution to the existing behavioural validation approaches is the distinction between the subject's behaviour and the subject/simulator performance which they characterise as validation **"man-in-the-loop"** simulation. Also the distinction they make between the controlled experimental and the uncontrolled observational conditions under which the real road experiment takes place, when referring to the comparison of performance differences between the two environments (simulator and real road). They suggested that when simulated data are compared to uncontrolled observational real road data, then this method "might be considered the highest form of validation". They also emphasised the critical issues of "operator motivation" and traffic scenarios in the simulator.

Allen et al (1991) stated that validity can be checked at several levels including the sensory cueing response to control inputs and measured task performance compared with real road data. The simulator cueing response to control inputs can be broken down into the component reactions of the vehicle dynamics and the cueing device (i.e. visual display, motion platform and control load system). Measured task performance can be compared with real world performance at several level ranging from the dynamic response of the man machine system, to overall system performance measures such accident rates. Validation of simulator components should pay particular attention to this critical aspect of simulator performance (the visual and motion display pathways show delay compensation that is intended to counteract response delays and lags in the display generator and motion base) and verify the effectiveness of compensation techniques. Cueing delays are not typically a significant issue for other feedback modalities as they are not used directly for closed loop control.

Moraal (1981) and Alicandri et al (1986) approached the issue of the driving simulator validity as all the previously mentioned researchers (Mudd, McCormick, Brown, Blaauw, Allen et al and Bertollini et al). The difference is that they used the terminology **functional** and **face validity** in terms of behavioural and physical validity.

Finally, Boulanger and Chevennement, (1995) They state that "an absolute and global validity has no meaning" thus, it is important to consider validity in relation with specific users' fields of work. For example, engineers who study accident prevention are interested in extreme driving conditions whereas ergonomists are primarily interested in common driving conditions. They are not interested in the same meaning of "driving activity" (in terms of mental and physical occupancy) and their validity needs are totally different. Therefore, they based their theory on "the functional declination of the objective of the Driving Simulator which is first to simulate a Driving Activity...". They emphasise the need for criteria which will allow the conclusion whether a specific driving simulator is suitable or not for a particular experiment and stress the fact that the specifications of that driving simulator must be known in detail (e.g. fixed-base or moving-base, complexity of the scene, resolution, number of projectors). Their conclusion is that "the gained experience along the time must be capitalised [adapting] either the simulator or the protocols (methods)". They approached the validity of driving simulators problem by distinguishing it to the **"analytical"** and **"experimental" approach**.

The analytical approach divides the "Driving Activity" into three main levels. The first one includes the functional objectives of a particular experiment (e.g. data collection and analysis for a particular experiment). The second level relates to the simulator context (i.e. its description and specification for that experiment) and the third to the technical parameters of the simulator (e.g. visual, sound systems) which are directly related to the particular experiment. According to them the three most important criteria and/or specifications for a particular experiment are: a) high complexity -or not- of the scene display (traffic, textures) (Padmos, 1992); b) necessity -or not- of a large visual restitution or a rear view and c) large range -or not- of dynamic cues (which requires -or not- a motion base) (Benson et al 1989). All the advantages and disadvantages of the simulator are available to the user, therefore s/he is aware of each subsystem's capability and the simulator team can propose the best configuration. The driving activity must be well known and well detailed, thus it is necessary to regularly adjust the method to avoid a too long procedure. The disadvantage is that the experimenters are not always able to define exactly what is useful for the experiment.

The experimental approach is additional to the analytical approach because it uses the experimenters' expertise and thus the time required to verify if the simulator is corresponding to the schedule of conditions is minimised. It focuses on finding general indicators which are suitable for a specific experiment, thus is using the skills of the experimenters (e.g. if a traffic engineer considers rumble strips as a speed-reducing measure on the real road, then the simulator validity should be tested using this parameter). The disadvantages is the conclusion of a validation test. Either results are correct and the user knows that he could use the simulator if he has to lead similar experiment in the future. But he knows that it is valid only in the same conditions. Or the results are not acceptable and it is difficult to deduce which subsystem of the simulator have to be improved. There are no obvious links between the indicators and the technical parameters of the simulator. In other worlds, this method is easier to develop but does not facilitate the analysis. The solution is to link both in establishing the relation between the functional objectives and the experimental indicators.

All the above mentioned approaches for the validation of driving simulators are summarised in Table 4-1: Summary of driving simulator validation approaches.

Researchers	Approaches		
Mudd (1968), McCormick (1970), Brown (1975), Blaauw (1982)	**A) Behavioural** 1. comparison of two systems during identical tasks and circumstances in terms of system performance and/or driver behaviour 2. measurement of physical and/or mental workload 3. subjective criteria from drivers (questionnaires) 4. evaluation of transfer effects		**B) Physical** 1. comparison of the simulated and the actual vehicle (e.g. geometry of control and their response characteristics)
Bertollini et al (1986)	**Closed-loop** 1. performance and performance trends 2. subjective ratings correspond		**Open-loop** 1. simulated and actual vehicle response characteristics
Allen et al (1991)	**Operator behaviour** 1. operator's subjective reaction (simulator fidelity) 2. operator's objective behaviour (perceptual and control responses, judgements and decision making)	**Operator/simulator performance** 1. transient response to isolated events and mean and variance response to random inputs 2. demonstration of transfer of training to real world performance	**Verification of simulator component response characteristics** 1. simulated vehicle response behaviour (i.e. vehicle dynamics or equation of motion) 2. response behaviour of the various simulator cueing devices (e.g. visual, motion, auditory displays)
Moraal (1981), Alicandri et al (1986)	**Functional** 1. comparison of performances between the simulator and the real world		**Face** 1. physical correspondence between the simulator and the real vehicle
Boulanger & Chevennement (1995)	**Analytical**		**Experimental**

Table 4-1: Summary of driving simulator validation approaches

4.1.1. Comparison of validation approaches

The review of the driving simulator approaches showed that all the previously mentioned researchers (Mudd, 1968; McCormick, 1970; Brown, 1975; Blaauw, 1982; Allen et al, 1991; Bertollini et al, 1986; Moraal, 1981; Alicandri et al, 1986 and Boulanger and Chevennement, 1995) agree that a simulator has to be validated both behaviourally (subject driving in the simulator and the real road) and physically (simulator versus real vehicle). The majority of them assume that drivers behave in the simulator as they behave on the real road when driving under similar conditions and also that the accuracy with which the real road data are measured and recorded is the same as the accuracy with which the simulated data are recorded for similar driving conditions.

The Allen et al (1991) approach questioned the above assumptions and addressed the problem by considering the "operator motivation" and the different ways of measuring and recording data on the real road. When driving a simulator, which inherently provides a safe environment, there is lack of time pressure and of the feeling of being under risk. This means that the driving speed in the simulator and the accuracy of driving the vehicle (e.g. lateral position, overtaking manoeuvres) may vary significantly from the real road driving. Also, traffic scenarios can have a strong influence on the "realism" of the simulation and thus some influence on subject motivation too. They suggested that "incentives must be set up creatively in order to minimise game playing and generally encourage speed/accuracy trade-offs consistent with real world conditions" and in particular if subjects driving the simulator are motivated by using monetary values then their behaviour may be closer to their real world behaviour.

They also examined the methods of collecting real road data to compare them later with simulated data. The accuracy of the real road collected data can vary according to the method used. Data can be recorded under controlled experimental conditions (e.g. subjects drive an instrumented vehicle on a real road or a test track in presence of the experimenter) or uncontrolled observational conditions (e.g. genuine road users behaviour can be recorded by using inductive loops or video cameras). According to the author, this element makes also a distinction between the use of instrumented vehicles and test tracks and the use of genuine road users data. Most of the validation studies have used instrumented vehicles, either on a real road or on a test track. This introduces a number of problems such as: a) both the instrumented vehicle and the simulator are part of an artificial environment. Drivers in both conditions are aware of the fact that they are not driving their own car and that the experimenter and the technician are closely watch their driving behaviour -in the instrumented vehicle case it is even worse because both of them are located inside the vehicle; b) the use of a test track in combination with an instrumented vehicle give data far closer to the data obtained from a simulator than from genuine road users. To the author's knowledge there are no studies comparing data taken from instrumented vehicles and genuine real road data to investigate the influence of the experimenters inside the vehicle and the influence of driving a new vehicle on a test track with no other road users to driver's behaviour. Evans (1991) defining driver behaviour states that "as driver behaviour indicates what the driver actually does, it cannot be investigated in laboratory, simulator or instrumented vehicle studies". This suggests that simulators and instrumented vehicles belong to the same category when apply to driver behaviour and any results obtained from an instrumented vehicle can be so uncertain as if they were obtained from a driving simulator.

Although some of the reviewed approaches take into account subjects' opinion about simulator realism using questionnaires, it seems that there is no standardised method for their design (i.e. the questions used should be the same for all the different type of simulators) and it is questionable if overall subjects' opinion is really taken into consideration for the further development and improvement of the simulator. It is also apparent that there is no link between behavioural and physical correspondence. Most driving simulators are validated either "behaviourally" or "physically". However, driving simulators are continuously upgraded and new features are added to them (from rear projectors, to complicated traffic scenarios and intelligent traffic), but experiments are not repeated with the upgraded version, therefore it is impossible to find if this upgrade significantly improved the simulator on not.

Both behavioural and physical correspondence are important for the successful validation of a simulator and have been mentioned in all validation simulator approaches. On the other hand only Allen et al (1991) mentions the **cognitive and/or perception correspondence**. The lack of cognitive and perceptual correspondence can also be the reason why it is still not known which of the real-road driving cues are of the greatest importance. However assuming that we do know, we implement those cues on the simulated driving and sometimes we end up in false or even contradictory results. For example it is assumed that the sight distance on real road and the simulator are the same and also that the perception of distance on the real road (which is a three dimension field) is the same as in the simulator (which is a two dimensional field). Is our assumption correct? Since the answer is not exactly known yet, the author, who is currently undertaking the Leeds Advanced Driving Simulator validation study, decided to take into consideration not only the behavioural and physical correspondence but also the perceptual correspondence.

According to Michon (1985), the unsatisfactory cognitive approach to the real driving task from most of the driver behaviour models could be due to lack of new, "striking" ideas about this topic and thus lack of money to support this type of research. It could also be attributed to the fact that a simulator validation study is a multi-disciplinary task that requires the best understanding and communication skills between all the different disciplines which work for its successful completion, something which does not happen in reality.

4.2. DRIVING SIMULATOR METHODOLOGIES FOR ASSESSING THE VALIDITY

Although numerous validation theories and approaches have been proposed since the conception of the simulators (either flight and/or driving), there is only methodology, according to the author's opinion, in terms of describing in detail all the steps to be followed in order to validate a simulator, the one proposed by Leonard and Wierwille in 1975.

Leonard and Wierwille (1975) proposed a methodology for assessing both the physical and behavioural validity of a driving simulator. They found that "the concept of performance validation is both α-level and sample size dependent, indicating that careful preliminary consideration should be given to the size of experiment to be performed".

Their validation approach can be described in the following steps.

Step one: Define the validation approach

> The validation approach is to "adjust the simulator experimental conditions to obtain matching measure values between full-scale and simulation".

Step two: Define the validation objectives

"Determine whether an absolute matching of driver and driver/vehicle responses will result in an effective method for validating a driving simulator".

Step three: Define the independent variables

The independent variables are the adjustable parameters. Each adjustment e.g. roll, yaw, roll damping, lateral translation gains and steering sensitivity in the simulator may affect the subject's responses.

Step four: Define the dependent variables

These variables must be measures which theoretically can be obtained both in the simulator and on the test vehicle (or "full-scale" vehicle). These can include average steering wheel reversals over time, RMS lateral acceleration and average velocity standard deviation.

Step five: Define the type of statistical test

There are two different type of tests that can be used
1. the high power statistical test
The number of subjects (N) is large and the α-level is low. This means that the results may show statistical differences between the two conditions (simulator and real road) when in fact the actual difference is not of practical value and
2. the low power statistical test
The number of subjects (N) is small and the α-level is high. This means that the results may not find statistically significant differences between the two conditions (simulator and real road) when in fact the actual difference is crucial in a practical sense (Ellis, 1967).

Step six: Analysis of the results

This is the last step of their methodology where the results from both conditions are compared and analysed (assuming that the real road data have already been collected). The analysis can be as follow:

a) Detect and remove the simulated data which prove to be significantly different from the real-road data.
   i) As a preliminary analysis use the analysis of variance to test if there are significant differences on each of the dependent variables.
   ii) Use the "t" or "F" or Dunnetts' test to examine the nature of these significant differences. When comparing each adjustable parameter of the simulator and the test vehicle, the use of "t" or "F" test can be tedious.

b) Determine which of the remaining non significant conditions produces the best matching data to the full-scale system.
   i) use correlation analysis only if the same subjects were tested in both conditions or
   ii) use confidence interval error term in any other case. It can be applied as follows:
      a) evaluate the confidence limits (95 percent) for both conditions for all settings over all performance measures and

b) determine the error by combining an upper and lower confidence limit deviation between the simulator condition and the appropriate control (adjustable parameter). Any of the three following different equations can be used to determine the error. Each equation will produce different absolute results, but the relative results should be the same in most cases.

$$\text{C.I. error}_1 = \sqrt{e_u^2 + e_l^2}$$

$$\text{C.I. error}_2 = \sqrt{|e_u| + |e_l|}$$

$$\text{C.I. error}_3 = |e_u| + |e_l|$$

where C.I.= confidence interval
e_u= difference in upper confidence limits
e_l= difference in lower confidence limits

c) After the correlation coefficients or confidence levels are determined they can be grouped for purposed of determining the best simulator condition for the performance measures

i) for correlation coefficients

a) either the largest correlation coefficient or the smallest error term for an individual performance measure and

b) for more than one by summing the correlation coefficient values for each condition over the appropriate performance measures).

ii) for confidence intervals

a) the same procedure can be followed but normalising is required over each performance measure.

They concluded by suggesting five criteria for a successful validation study:

A. *"The simulator must possess good fidelity in those aspects corresponding to the measures taken.*
B. *The simulator must have the capability of parameter adjustment.*
C. *A sufficient number of properly selected independent variables and corresponding settings must be employed.*
D. *Performance data must be obtainable for the standard full-scale vehicle and for each adjustment of the simulator and*
E. *Accepted methods of experimental design must be used to insure unbiased data and correct conclusions regarding validity".*

4.3. DRIVING SIMULATOR VALIDATION CRITERIA

Whichever approach or methodology has been used for validating a simulator the final issue is the interpretation of the results after the comparison of the two environments. If the results are primarily concerned with driver behaviour and transferability of driving behaviour to real world then we are referring to the internal and external validity criteria; if they are primarily concerned with driver performance and performance differences between the two environments then we are referring to the relative and absolute validity criteria.

Internal validity can be used to recognise possible apparent relationship between a manipulation (e.g. using speed limiters) and an obtained effect (speed reduction). It can be achieved if there are no alternative explanations for an obtained effect but can be lost if driver

behaviour is specifically affected by the limitations of a driving simulator. Other problems can be the limited resolution of a computer-generated image, the delay until vehicle position and images are updated and a limited horizontal field of view. External validity can be achieved if the results (e.g. driving behaviour) obtained from a particular simulated experiment (e.g. using a specific set of subjects driving in a particular traffic scenario on a particular road environment during a specific period of time) can be generalised and used for driving on real roads. Problems can be caused by careless choice of road environment (e.g. road type) or subject selection (e.g. amount of driving experience), motivation and mental and physical condition (fatigue of subjects). It may related to the design of an experiment on the basis of a specific research question.

In addition to the internal and external validity criterion, the relative and absolute criterion can be used. Relative validity, a qualitative criterion, is achieved when the performance differences of a subject driving on the real road and the simulator under similar conditions are of the same order and direction. Absolute validity, a quantitative criterion, is achieved if the numerical values of these performance differences are about equal or the same. Most researchers have used the absolute and relative criterion for validating their driving simulators (Blaauw, 1982; Harms, 1993; Alm, 1995; Kaptein et al, 1995; Reed and Green, 1995). This criterion applies mainly when investigating driving activities/tasks on the control level when comparison of variables between the two environments are less complicated to be made. On the other hand, when investigating driver's behaviour on the tactical and strategic level (due to a number of interactions between the drivers and the other road users), the comparison of results from the two environments is more complex. Therefore, if the criterion of absolute and relative validity cannot be used then the internal validity criterion should be used.

Although relative and absolute validity are supposed to be the criteria to assess driving simulator validity, these criteria are rather general and "relative" from simulator to simulator. A number of researchers have claimed that their driving simulator is valid on "absolute" terms but the thresholds (numerical values) they used when comparing the results from the two environments are not known. Thus this characterisation is arbitrary. It is therefore necessary to adopt a common approach, or strategy, to tackle the problem of validating a driving simulator.

5. REVIEW OF EARLIER AND RECENT BEHAVIOURAL VALIDATION STUDIES

The term "behavioural validity" of a driving simulator is defined as the comparison of driving performance indices from a particular experiment on real road with indices from an experiment in a driving simulator which is as close as it can be to the real experiment, i.e. the same road network, the same type of car, the same drivers, the same road environment and other traffic.

The issue of behavioural validity have not been addressed before 1975 for driving simulators because they were still in the developing stage but it was already a problem for the aircraft simulators. However validity had been addressed in terms of fidelity and its effects on transfer of training (Mudd, 1968; Blaiwes et al, 1973; Caro, 1973; Provenmire and Roscoe, 1973; Valverde, 1973; Williges et al, 1973). The first driving simulator validation studies examined more the physical rather than the behavioural correspondence of the simulator to the real world.

A number of behavioural validation studies have been examined here. For the early studies they are not exactly known all the technical characteristics of the simulators that they were used, or the type of statistical analysis was used and great detail about how the simulated and real road experiment have been conducted. The later validation studies have been described in greater detail and all the technical characteristics of the simulators used can be found in the relevant papers as well as in a survey about the most known driving simulators around the world by Blana (1996). All the details of the validation studies -both for the simulated and the real road experiments- are described in Table 9-1, Appendix A.

Before the review of any driving simulator behavioural validation study it was considered necessary from the author to give the definitions of driver behaviour and driver performance and the distinction between these two definitions as well as the definition of the driver behaviour levels as this terminology will be widely used in the following paragraphs.

5.1. DRIVER PERFORMANCE AND DRIVER BEHAVIOUR

"Driver performance... refers to the drivers' perceptual and motor skills, or what the driver *can* do whereas driver behaviour refers to what the driver in fact *does* do" (Evans, 1991 p.133). Driver performance focuses on capabilities and skills and can be investigated by many methods, including laboratory tests, simulator experiments, tests using instrumented vehicles and observations of actual traffic. On the other hand, driver behaviour cannot be investigated in laboratory, simulator or instrumented vehicles studies. Therefore, information on driver behaviour tends to be more uncertain than that of driver performance.

The distinction between driver behaviour and driver performance is one of the most central concepts in traffic safety because according to Näätänen and Summala (1976) driving is a "self-paced" task. In other words, drivers choose their own desired levels of task difficulty. The driving task is a closed-loop compensatory feedback control process, meaning that the driver makes inputs (to the steering wheel, brake and accelerator pedal), receives feedback by monitoring the results of the inputs, and in response to the results, makes additional inputs; an open loop process is one, such as throwing a baseball, in which once the process is initiated no corrections are possible based on later knowledge about the trajectory (Evans, 1976 p.109).

Relative to driving simulators, Crawford's (1961) statement thirty six years ago, that "it has proved extremely difficult to define what is meant by driving performance and to develop adequate techniques of measuring it" still stands today although according to the author's opinion Crawford is referred to driver behaviour rather than driver performance. Nilsson (1989) has also pointed out it is not exactly known which of these cues are really necessary and essential for a successful representation of the real road environment in the simulator. Nilsson defined the traffic system for traffic safety purposes as accidents, physiological measurements and driving performance. The way these measures are actually chosen in a study are strongly dependent on the hypothesis to be tested in that specific study and can be any variable that is available in the simulator model. Physiological measures include the monitoring of physical and mental stress of the body from the environment and the driving tasks, as well as the level of arousal (e.g. pulse rate, blood pressure etc.). Other miscellaneous measures include questionnaires and interviews to detect the participants' subjective opinions and evaluations concerning the tested tasks, conditions, etc. Driving performance includes, most frequently, forward speed, lateral vehicle position and different stimulus-induced reaction times. Less frequently lateral or longitudinal accelerations, steering wheel angle and steering wheel torque are measured (Nilsson, 1989).

Limitations in simulator validity are directly related to the cues that a specific simulator provides. On one hand drivers rarely use all the available cues to perform a task. Thus according to Flexman and Stark (1987), it is not always necessary to provide in the simulator identical cues to those of real life. On the other hand it is assumed that all types of real road environment cues (e.g. visual information, sound, self-motion) are provided more or less to the simulator. Thus it can also be assumed that the results observed in the simulator can be successfully compared to the results obtained from real life.

This problem of identifying which cues are the most important for the real-world and the simulated driving could be attributed to our limited knowledge of how drivers perceive and understand the road environment and how exactly they behave and interact with other road users while driving. Therefore, a first approach to the solution of this problem could be an attempt to define driver behaviour and the levels that a driver progresses through (consciously or unconsciously) when s/he drives from a point A to a point B.

5.2. DRIVER BEHAVIOUR LEVELS

According to Janssen (1979) and adopted by Michon (1985), driver's behaviour can be described by three levels: the strategic, the tactical and the control. Each level is defined by different action patterns and a different "preview" which is the time in which the events, that are correlated with and dependent on the behaviour in the actual situation, will take place. Figure 5.1 gives diagrammaticaly the three driver behaviour levels.

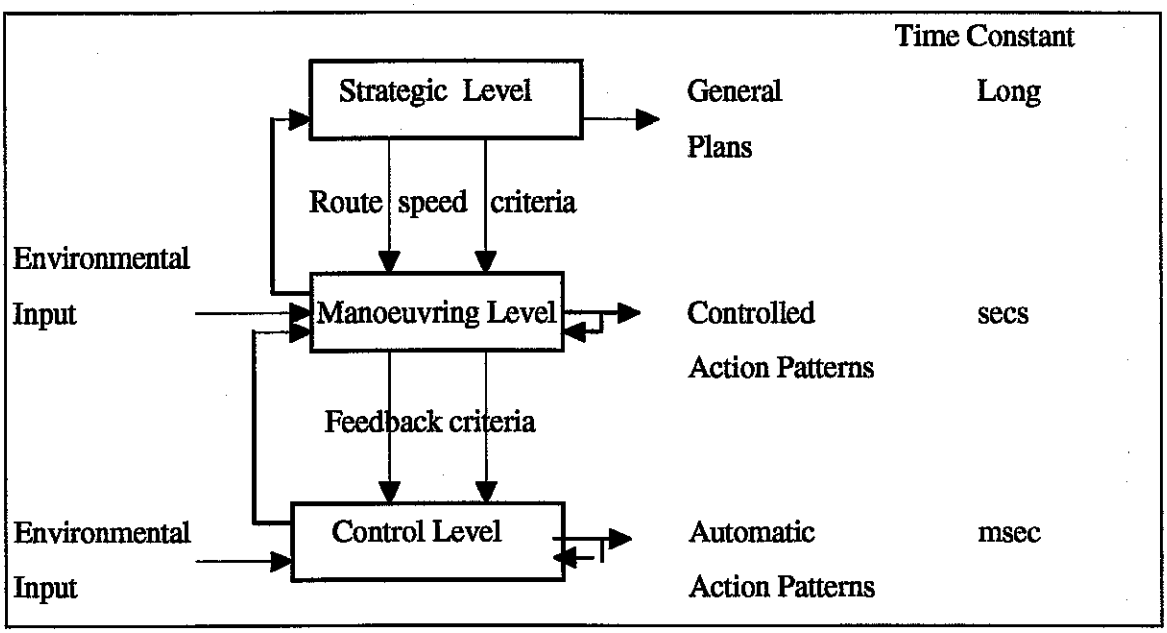

Figure 5.1 The hierarchical structure of the road user task (after Janssen, 1979).

The strategic level is mainly related to the process of route planning, and following of a route using various means of route information. It includes the determination of trip goals, route and modal choice, plus an evaluation of the costs and risks involved. Plans derive further from general considerations about transport and mobility and also from concurrent factors such as aesthetic satisfaction and comfort. At this level, the preview can be as long as the whole drive. The driver is fully aware of the different tasks. Usually in-vehicle navigation systems are tested in the simulator at this level.

The tactical level is mainly characterised by the manoeuvring behaviour (e.g. overtaking, crossing and turning, obstacle manoeuvring and gap acceptance). These patterns must meet the criteria derived from general goals set at the strategic level. Conversely these goals may occasionally be adapted to fit the outcome of certain manoeuvres. In this case, the preview is of the order of seconds to a few minutes. The assimilation of information, and decision-making, are more conscious than at the control level. Simplified in-vehicle information systems, mobile phones, speed limiters, automatic-cruise controllers are tested in the simulator at this level. (the steering-wheel movements demonstrate a difference in control tactics).

The last level, the control level, defines the automatic action patterns. The tasks which are situated here have the purpose of adjusting the position of the vehicle on the road both in longitudinal and lateral directions. Steering of the vehicle and steering it on the road and choosing speeds and gears are the relevant tasks. Two important things about the control level are that the "preview" is of the order of a few seconds or less and the different tasks are accomplished in an automatic way: the driver is hardly aware of the visual information s/he assimilates and of the way in which this information results in decisions and actions. Traffic calming measures, new tunnel design, impaired driving and experiments which are directly related to the longitudinal and lateral control of the vehicle are tested in the simulator at this level.

In order to be able to perform the different tasks at each of these three levels, the driver needs to get information about the conditions of the surroundings. If the preview time is shorter, the need to update the information available is more frequent. This means that at the micro level, an almost continuous information stream is necessary, while at the strategic level, the information can be spread over a longer timescale and it doesn't have to be continuous.

How are these driver behaviour levels related to the behavioural validation of driving simulators? When conducting experiments on research driving simulators, a combination of different action patterns out of the three different driver behaviour levels is often used to describe the experiment and more often these are the control patterns.

5.3. EARLY BEHAVIOURAL VALIDATION STUDIES

Barrett et al (1965) evaluated the equipment fidelity of a driving simulator using acceleration, braking, turning radius and emergency behaviour as dependent variables. Not all of these data could be compared in both real road and simulated conditions. The real road data were either standard references or from a test vehicle. The biggest problem for this study was simulator sickness (64 percent). This study is cited in Leonard and Wierwille (1975) who did not proved any results for the Barrett et al validation study.

Wheaton et al (1966) investigated the validity of a part-task driving simulator in terms of corresponding data (both absolute and relative), subjective data for face validity and engineering evaluation data of the simulator's display and control fidelity. They used overtaking/passing manoeuvre and car following as dependent variables and different levels of ambient illumination, overtake rates and different taillight configurations as independent variables. The results showed low absolute correspondence (using comparison of means and the null hypothesis) but high relative correspondence (using an analysis of relative trends). The subjective data showed that in general the face validity of the simulator was good.

Wojcik and Weir (1970) compared simple driving manoeuvres in a scale model simulator (with a black and white 40 degrees image and a mock up on a rolling-road so that speed-dependent rear wheel vibrations were fed back to the drivers) and in the field including overtaking, driving on a curved road, lane keeping with side wind and following a lead vehicle. Their results showed the same relative changes in the simulator and in the field, which suggest relative validity of their simple simulator for the tasks that were tested.

Breda et al (1972) compared driver performance as a function of different types of route guidance systems on real road and in a simulator. They implemented two methods to examine the validity of their simulator. The first method used the validity coefficient (product moment correlation coefficient) and the second one a sensitivity analysis to determine whether the independent variables affected both real road and simulator dependent variables in a similar manner. The results showed that the absolute measures of correlation were poor whereas the relative measurement analysis produced much better results. They reported the problem of simulator sickness to several subjects and 7.5 percent of the subjects had to quit the experiment.

Leonard and Wierwille (1975) investigated the validity of a driving simulator using a completely different methodology (the one described in detail in paragraph 4.1.1). Their goal was to adjust the simulator "to match the response obtained from full-scale tests on an absolute basis". They used the adjustments of the simulator as the independent variables (random disturbance level and lateral display gain) and created eight adjustment conditions by varying these two independent variables. As dependent variables they used the response measures of the operator and the man-machine system (five performance measures for each condition). The real road data were collected using an instrumented vehicle. The results showed that "for each performance measure at least one simulator condition produced corresponding valid results".

Allen and O' Hanlon (1979) compared the effects of road marking contrast in their fixed-base simulator (with a black and white TV monitor and 67 degrees field of view and computer generated imagery) to those obtained from an instrumented car on the road. The results showed a clear decrease in variability of lateral position for increased road marking contrast, both in the simulator and in the field. For both conditions models were derived of the effect of marking contrast on variability in lateral position. A comparison showed that the simulator based model did not differ from the field based model. It was concluded that their simulator was valid for this type of research.

5.4. RECENT BEHAVIOURAL VALIDATION STUDIES

The definition of "recent behavioural validation studies" means validation studies in driving simulators after 1980 and generally after the development of the powerful workstations and computer-generated images subsystems of the simulators.

5.4.1. The TNO validation studies

Blaauw (1982) studied driving experience and tasks demand in a simulator and an instrumented vehicle on a real road. The simulator which was used for this experiment was the old model of the TNO driving simulator (a scale model). For the subjective performance of the simulator he used two questionnaires, one relative to task difficulty, required attention and monotony using a continuous scale 0(extremely unfavourable)-100(extremely favourable)

and one relative to the realism of the simulator (comparison of simulator and instrumented car using multiple choice questions) asking also the subjects for personal comments. The results of the first questionnaire are presented in Table 5-1.

The results showed that there was statistically significant difference ($p \leq 0.01$) for almost all opinions between the instrumented vehicle and the simulator for both groups of driving experience and task demand. As it can be seen, all drivers rated the simulator more unfavourable (task difficulty, required attention and monotony) compared to the instrumented vehicle with the exception of the longitudinal control (driving on a straight road with no other traffic). Experienced drivers judged more favourable ($p \leq 0.01$) than the inexperienced drivers with the exception of monotony.

Questionnaire Item	Instrumented car		Simulator	
	Inexperienced	Experienced	Inexperienced	Experienced
Task difficulty				
Overall	59	73	36	46
Lateral control	57	63	21	41
Longitudinal control	61	69	73	75
Required attention				
Overall	35	45	29	31
Lateral control	38	44	18	28
Longitudinal control	54	63	72	77
Monotony	55	38	9	9

Table 5-1 Mean ratings of drivers' opinions on various aspects of driving tasks

The results of the second questionnaire confirmed the findings of the first questionnaire, i.e. it was found greater task difficulty in the simulator. The subjects also commented about the monotony of the simulator due of lack of other traffic, road curvature and road signing. It should be noted that no one experienced motion sickness. The comparison of the real and simulated data showed that experienced drivers have a significantly smaller standard deviation of lateral position, yaw rate, and steering-wheel angle, but still it was greater in the simulator than in the real vehicle. Overall, there was "good absolute and relative validity for longitudinal vehicle control" but "lateral vehicle control offered good relative validity" due to "the larger variations in the lateral position in the simulator".

The experiment that Blaauw performed in 1982 was reconstructed by Kappé and Körteling (1995) using the new TNO driving simulator with its computer-generated image system. The results showed that this time there was no difference for inexperienced and experienced drivers in lane keeping behaviour in the simulator compared to real road. A possible reason for the invalidity of simulator in the first experiment could be the characteristics of the scale model.

Tenkink (1989) studied the effect of road width and obstacles on driving speed and steering behaviour on a test rack using the TNO instrumented vehicle (ICARUS). He later compared with a similar study in the TNO driving simulator (Tenkink, 1990). At the same time he conducted two more experiments in the simulator. One where the subjects drove with and without a auditory-verbal memory task and another one with and without a visual task, looking again at speed and steering behaviour. The aim of the studies was to verify that the two most important factor affecting drivers' speed choice are time gain and avoidance of

negative consequences such as accidents as he had concluded in his earlier respective literature review (Tenkink, 1988). The decrease of the negative consequences could happen with speed reduction and more accurate steering. for the validation experiment, the field trial subjects had to drive on lanes with different widths (2, 3 and 4m) and obstacles along both sides of the road (metal posts with red and white strips usually used for road works with height 1.35 and width: 0.25m). The effective widths of the road (the distance between the two obstacles) was 2.51, 3.51 and 4.51 respectively. The width of the car was 1.83m, therefore the corresponding margin between the car and the obstacles was 0.68, 1.68 and 2.68m respectively. For the simulated trial subjects had to drive again on lanes with different widths (2, 3 and 4m) and obstacles along both sides of the road (this time the metal posts were 1m height and 0.5m wide for prevented the simulated horizon to run through the posts). The effective widths this time were 2.40, 3.40 and 4.40 respectively. The width of the car was 1.72m, therefore the corresponding margin between the car and the obstacles was also 0.68, 1.68 and 2.68m respectively. Another lane with 2.01 road width was added to the simulated experiment in order to get an impression of the steering ability of the drivers. The driving instruction in both trials was to drive not too slow but without hitting any obstacle. The results showed that in both systems driving speed and variation in lateral position decreased when obstacles were placed nearer to the road. However, in the driving simulator subjects generally drove at slightly higher speed combined with a larger variation in lateral position.

Tenkink and Van der Horst (1990) studied the effect of road width and curve characteristics on driving speed in the simulator and compared the results with numerous earlier studies of other researchers on the real road. The results showed again relative validity of the simulator with regard to driving speed behaviour: both on the road and in the simulator driving speed reduced with decreasing road width and decreasing curve radius. There was no absolute validity: in the simulator higher speeds were chosen compared to on the road. For instance, in sharp curves, drivers chose speeds in the simulator that in the field would not have been possible. It should be noted here that this a different type of behavioural validation study because the experiment in the simulator is not exactly the same as the experiment on the real road, thus the results could be compared only in a qualitative and not quantitative way.

Hogema (1992) investigated the effect of a compensation technique for the delay in the visual display of a driving simulator. The real road experiment was a double lane change task on a test rack (ISO Technical Report, 1975) using the TNO instrumented vehicle (ICARUS) and two configurations of the visual display system of the simulator were user, with delay and with compensated delay (three vehicles). For this particular experiment, the visual angle of the simulator was 46 degrees. Subjects were asked to fill in a questionnaire which was done on a two-level rating scale and rated the difficulty of the manoeuvre. Each subject completed six (6) blocks, each block consisted of 3 runs: one at each of the 30 km/h, 45 km/h and 60 km/h speeds. It was found that there was no significant benefit from using the compensation technique in the driving simulator and the maximum safe speed was "not a meaningful quantity in the simulator". Overall, absolute validity have not been achieved for this lane change task because there were statistically significant differences between the simulator (with and without the compensation technique) and the instrumented vehicle. On the other hand, the results of the ratings and the steering reversal rate (SRR) were relative valid but not the results of the cone displacements.

Janssen et al (1991, 1992a, 1992b) and Van der Mede and Van Berkum (1993) conducted a number of simulator studies which later were compared to field studies. These studies were focused on the effect of variable message signs in route choice and driving behaviour. It was shown that a driver's choice behaviour was affected by both the individual cost of time loss and the degree that surrounding traffic follows the advice. There was absolute validity of

result obtained form the field study and the simulation: both average choice behaviour as well as the size and direction of effects were comparable (i.e. results showed that if time loss is important, drivers are least inclined to follow the example given by others, thus showing signs of an intention to outperform the system).

Van der Horst and Hoekstra (1993) investigated the perception of chevrons in fog in a driving simulator and Hogema et al (1993) investigated the same condition on a rural road. Comparison of results from the two studies showed that there was relative validity: both on the road and in the simulator, the chevrons were of little use, since drivers proved not to be capable to concurrently watch the chevron and paying sufficient attention to the driving task.

Kaptein et al (1996) investigated what visual information is important in a driving simulator by studying braking behaviour, including normal and hard braking, and Time-To-Collision (TTC). The simulated results were compared with an earlier field study on TTC (van der Horst, 1990). The real road experiment was replicated in the simulator using two levels of both scene complexity and field of view. The scene was a straight road section; either plain: textured road without lines but no road environment or complex: textured road with lines and road environment (houses, trees, delineator posts) and it was projected either with 40° or 120° horizontally field of view respectively. No comparative numerical values are given in Kaptein et al paper for the two types of braking, the scene complexity and the field of view but only figures. It can be seen that for low approaching speeds (30 km/h) most of the results had absolute validity for the hard braking condition. Results were different for higher approach speeds and normal braking. An important finding was the minimum TTC (TTCmin) that drivers accepted during the manoeuvre was constant over experimental conditions in the field study, whereas in the simulator it increased with approach speed and with normal compared to hard braking. Apparently, in the simulator subjects need a relatively large safety margin if decisions have to be made at larger distances: at high approach speeds or with normal braking (using low accelerations, which implies starting to brake at a larger distance). They also found that with a simple scene the stopping distance decreased with field of view whereas with complex scene the stopping distance increases with field of view and "scene complexity showed not to be important" but "field of view is important during the braking manoeuvre".

5.4.1.1.Comparison of the TNO validation studies

The TNO behavioural validation studies which have been described in the above paragraphs, are summarised in the following table (Table 5-2). It can be seen that the results obtained from the different validation studies are rather contradicting. This means, that the simulator is claimed to be absolute or only relative valid for the same driver performance measures (e.g. for speed) which suggests that validity can only apply to a study similar to the one already conducted using the same measures of driving performance but not to any other type of study using the same measures and only using this particular driving simulator. This means that there can be no validity generalisation for the TNO driving simulator.

The available data from the TNO behavioural validation studies do not allow further analysis and/or comparison of the findings for various reasons such as: most of the detailed reports including raw data and test protocols (specifications) are produced in Dutch and only the abstracts or a synopsis of the experiment is translated into English; the TNO driving simulator is constantly updated and improved during all these experiments and although these improvements are always stated in the reports, generally, the effect of them to the studies is not exactly known.

TNO validation studies*	Dependent variables	Results relative to the validation criteria	
		absolute	relative
Blaauw (1982) **	speed and lateral position on straights	***YES*** for speed	***YES*** for lateral position
Tenkink (1989,1990)	speed and SD of lateral position		***YES*** for speed and SD of lateral position
Tenkink and Van der Horst (1990)	speed on curves	***NO*** overall	***YES*** for speed
Hogema (1992)	ratings, SRR, cone displacement	***NO*** overall	***YES*** for ratings and SRR
Van der Horst and Hoekstra (1993) and Hogema et al (1993)	chevrons perception in fog		***YES***
Janssen et al (1991, 1992a,b) and Van der Mede and Van Berkum (1993)	average choice behaviour by variable message signs	***YES*** overall	
Kaptein et al (1996)	TTCmin, TTCbr, ACCmin and DISTbr by hard and normal braking and low and high speeds	***YES*** for DISTbr and TTCbr but for low speeds and hard braking	***YES*** for DISTbr and TTCbr for either type of braking or speed, ***NO*** for TTCmin and ACCmin for either type of braking or speed

* all the TNO validation studies have been conducted in relatively different configurations of the driving simulator. For detailed information, the readers could address directly either to the technical papers of TNO or the authors of these papers

** this study used the old version of the TNO driving simulator which was a scale model

Table 5-2 Summary of the TNO driving simulator behavioural validation studies

5.4.2. The VTI validation studies

The behavioural validity of the VTI moving-base driving simulator has been examined by Harms (1993), Alm (1995) and Harms et al (1996). The results of these validation studies are presented in Table 5-3.

Harms (1993) tested the simulator validity using speed, lateral position as independent variables for the two conditions (real road and simulated). At that time the VTI simulator animation software was generic (only the road and plain scenery could be simulated) as well as the traffic modelling software (no other traffic could be simulated). She found both relative and absolute validity of the simulator for speed but only relative validity for lateral position. "Considering the between subjects variation in driving speed as an error term in the analysis of variance the factors driving condition and driving session accounted for only 15 percent (r^2=.15) of the variance and neither the effect of driving condition ($F(1,36)=3.67$, $p>0.06$) nor the effect of training (driving session) ($F(2,36)=1.44$, $p>0.25$) were significant. Driving session and driving condition accounted for about 50 percent (r^2=.52) of the variation in lateral position, but only the effect of driving condition was found significant ($F(1,32)=741.44$, $p>0.001$). The product moment correlation for driving speed in 5m intervals of the road within driving conditions was 0.97 and between driving conditions 0.87

(average). For the lateral position and within driving conditions the mean correlation was 0.925 and for between driving conditions 0.49 which demonstrates a less consistent driving pattern between the two driving conditions than within each condition. Her results with regard to lateral position are in accordance with Blaauw's (1982) findings. She suggested that this problem can be due to the absence of other traffic, or that the subjects use other visual cues for their lateral control in a driving simulator than during field driving.

Alm (1995) using the updated version of the VTI driving simulator (complex road environment and other traffic could be simulated), repeated Harms (1993) validation study. A new element was added here. He compared driving simulator experimental data with and without kinaesthetic feedback. He used speed and speed variation and lateral position and lateral position variation as performance measures and he used the NASA-TLX test (Hart and Staveland, 1988) to measure the mental workload after driving on a driving simulator and a questionnaire to measure the subjective realism of the simulated road on the VTI simulator with the moving system on and off. The questionnaire included the following questions: 1) How realistic was it to drive the simulator? 2) How realistic was it to drive the simulator on straight parts of the road? 3) How realistic was it to drive the simulator on curvy parts of the road? 4) Did you experience any type of nausea during the simulator trip. He used a seven point scale (1=not at all and 7=very much) to assess the data.

The results showed no statistically significant differences on average speed and lateral position for both environments. He concluded that the moving-base system is better when driving in curves, minimises the nausea effects from the simulated road environment and helps the driver to keep the car on a steady course on the road. The moving system had no effect in the variation in speed level. Still there was statistically significant increase in speed variation ($F(2,48)=10.24$, $p=.0002$) and lateral position variation ($F(2,48)=9.12$, $p=.0004$) compared to driving on real road. A Tukey HSD showed that the significant difference in lateral position was between the moving base off and the other two conditions (real road and moving system on). Another significant finding was that driving in the simulator produces higher mental workload compared to real car driving. Both relative and absolute validity of the simulator for speed and lateral position were found but there were statistically significant differences in speed variance between the two conditions and in lateral position variance with movement system off and between the two conditions with the movement system on.

The results of the NASA-TLX , using a one-way ANOVA, showed that driving in a simulator was more physically demanding ($F(1,34)=4.83$, $p=0.0348$), more effort demanding ($F(1,34)=10.06$, $p=0.0032$), and more frustrating ($F(1,34)=6.82$, $p=0.0133$), than driving on a similar real driving condition. Relative to the first question of the questionnaire for the subjective realism, there were no significant differences between the realism of the simulator when the moving system was on or off, still all the ratings had the same tendency that the realism is better when the moving system is on. Relative to the second question, driving on straight road sections with the moving system on or off wasn't very different (mean ratings 5.7 and 5.6 respectively), where driving on curves (question three) was rated more positive when the moving system was on than off (5.1 and 4.4 respectively). Finally (question four) the moving system off was rated as more nausea producing ($F(1,34)=2.53$, $p=0.1207$).

Harms et al (1996) in the latest validation study compared driver behaviour on a real and a simulated tunnel. Driving speed and lateral position were used as dependent variables, like in the two previous validation studies. The position of the tunnel wall (right or left side of the driver) and access to speedometer values of driving speed were used as independent variables. The results showed statistically significant difference in mean driving speed between the two systems (8 km/h higher in the simulator than on tunnel) whether or not there

was access to speedometer values. Possible reasons could be the width of the road and the fact that there was no other traffic (three lanes, one-way only). For the lateral position and the side of the tunnel wall, similar results were found. Subjects drove 40 cm closer to the right wall in both environments. Overall, access to speedometer and position of the tunnel wall both significantly affect driving speed and lateral position. Their overall conclusion was that "the presence of critical but unnoticed source of variance, influencing subjects speed and lateral position both in the field trials and simulator trials, may result in unreliable conclusion of behavioural validation studies".

5.4.2.1.Comparison of the three VTI behavioural validation studies

For the first two validation studies (Harms, 1993 & Alm, 1995) the same road network and the same road section of this network has been used for the data collection, i.e. a rural road, 7 m wide (3.50 lane width, one lane per direction). The third validation study (Harms et al, 1996) was totally different than the two previous and as road network they used a new-built tunnel not given to traffic yet. The same subjects were used in both real and simulated studies in each of the three validation studies. The results from the comparison of the field and simulator trial with regard to the mean speed and lateral position are given in Table 5-3 and Figure 5.1, Figure 5.2.

At this point the author would like to inform the reader that there is some inconsistency between the numerical values of the mean speeds of the three VTI validation studies (published in Harms, 1993; Alm, 1995 and Harms et al, 1996) and the figure which has been produced by using these mean speeds (published in Harms et al, 1996). Thus, if the reader compares the figure of this working paper (Figure 5.1) with the figure in Harms et al (1996), s/he will notice distinct differences. The author produced Figure 5.1 by using the numerical values published in the three VTI validation papers.

Comparison of the three VTI behavioural validation studies

	Road characteristics			Type of vehicles		Mean Speed (km/h)		Displacement (m)	
STUDIES	**Type of road**	**Lane width (m)**	**Speed limit (km/h)**	**Instr. vehicle**	**Simul. vehicle**	**Field trials**	**Sim. trials**	**Field trials**	**Sim. trials**
First study (Harms, 1993)	Single c/way	3.50	70-90	Volvo, 240 Sedan	Volvo, 240 Sedan	79	81.7	-0.03 (0.92)	0.20 (0.71)
Second study (Alm, 1995)	Single c/way	3.50	70-90	SAAB 9000	SAAB 9000	83.9	84.02[x]	0.15 (0.73)	0.08 (0.78)
Third study (Harms et al, 1996)	Tunnel (3 lanes)*	3.25	70	SAAB 9000	SAAB 9000**	73.4	81.0[+]	0.04	-0.09

[x] This is the mean driving speed with the moving system on (with the moving system off is 85.07 km/h)
* the tunnel wall was either on the left or the right side of the driver (i.e. the driver had to driver either closer to the left or the right tunnel wall)
** some of the dynamic properties of the real SAAB 9000 were actually simulated (it wasn't the case in the two previous studies)
[+] this is the mean driving speed with access to the speedometer (without the speedometer was 84.7 km/h)
(...) the parentheses give the values of lateral position of the left back wheel of the vehicle to the centreline

Table 5-3 Mean driving speed and displacement of the centre of the car from the centre of the lane. Positive values indicate driving closer to the centre of the road, negative values indicate driving closer to the road edge.

Source: Part of data have been adapted from Table 1, Harms et al (1996)

Mean driving speed

Comparing the results of the three studies for the mean driving speed from Table 5-3 and Figure 5.1, it can be seen that there is some consistency in the speed the subjects choose to drive in the simulator although always higher than on the real road but there is no consistency in their driving speed on the real road (the speed limit was the same for all three studies, equal to 70 km/h). This inconsistency is really inexplicable for the first two studies where the road network was exactly the same and the only differences were the opposing traffic (one could think that the opposing traffic could affect driving speed in a negative way, i.e. minimising it) and the road environment (more complicated and textured the second time). The results from the first and third study (totally different road network but still no opposing traffic for both studies) are more comparable than the results of the first and second study (exactly the same road network). This result is really ambiguous because in the first study the road image was really limited (a plain grey road with white lines but no other road environment or road texture) and the dynamics of the simulated vehicle was different from the instrumented vehicle where in the third study both road image and the dynamics of the simulated vehicle were really high. It seems that the subjects choose a driving speed in the simulator irrespective of the road type (lane width, speed limit, road curvature, roadside obstacles). This finding cannot be supported from real road data where road type plays a important role in driving speed.

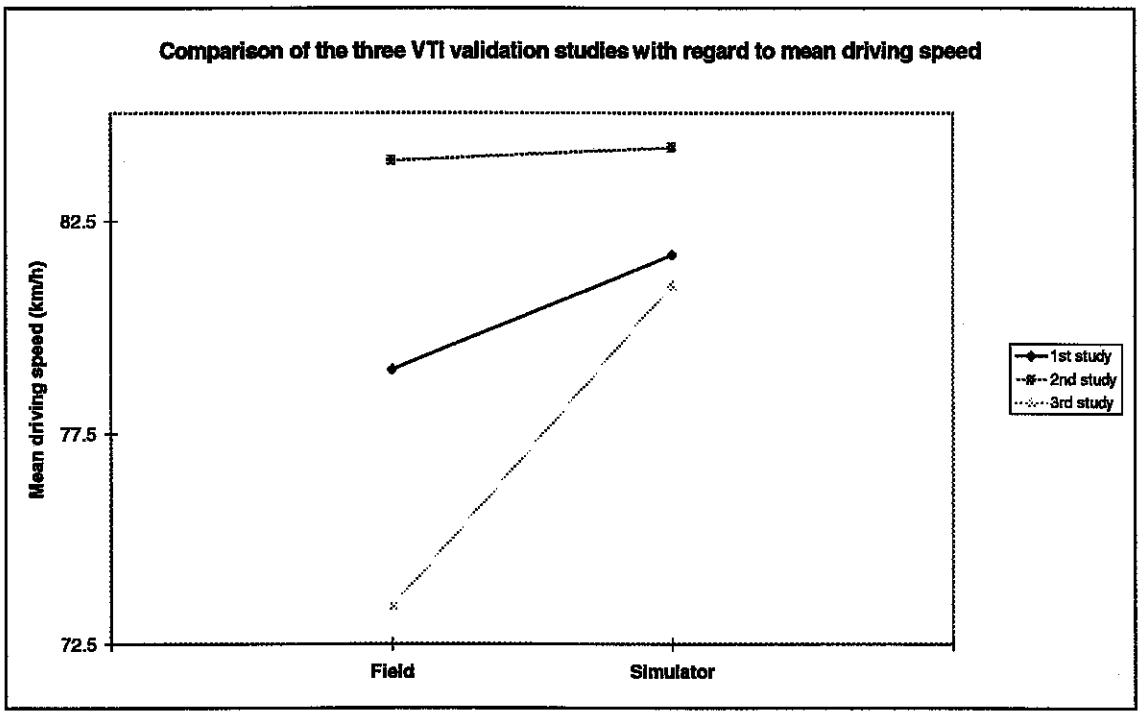

Figure 5.1 Comparison of the three VTI validation studies with regard to mean driving speed.

Mean lateral position

The comparison of the lateral position was impossible because different size instrumented vehicles (by 8 cm) and different types of roads and lane widths have been used in the three studies. Therefore, Harms et al (1996) had to define a new measure which could compare the

lateral position of all three studies: "the lateral deviation of the car-centre form the centre of the driving lane" or displacement. For the third study, it is not exactly known is if the displacement represents mean values of displacement for the tunnel wall been right or left of the driver.

From Table 5-3 and Figure 5.2 it can be seen that there is not such obvious consistency in the lateral control of the vehicle, either for the real or the simulated environment, like for the mean driving speed between the three studies. In the first study, subjects drove closer to the road edge in the field trial, in the second one there was no significant difference between field and simulated trials whereas in the third one they drove closer to the road edge (tunnel walls) in the simulated trial. According to Harms et al (1996) these differences indicate that "other factors than face-validity in simulator trials affected the subjects lateral position". For the first study, since subjects driving on the real road may expected opposing traffic they drove closer to the road edge where in the simulator, there was no opposing traffic. For the second study, the problem of no opposing traffic had been corrected, thus no differences observed. On the third study, subjects drove closer to the tunnel wall in the simulator indicating that the wall does not impose the same risk as in real life.

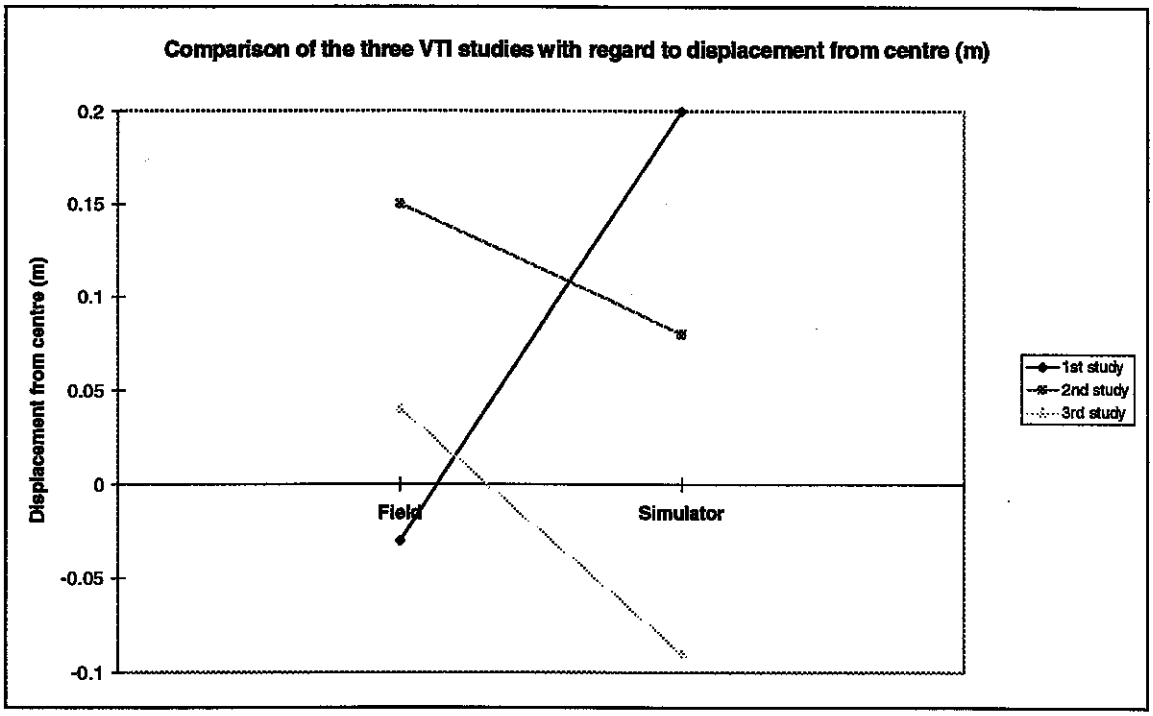

Figure 5.2 Comparison of the three VTI validation studies with regard to displacement from centre

Face validity
Although face validity was increased (especially in the third study), it didn't significantly improved the VTI simulator validity which "suggested that other factors than the face-validity of driving simulators may affect the subjects' driving behaviour" (Harms et al, 1996).

5.4.3. The INRETS driving simulator validation study

Malaterre (1995) assessed the validity of the INRETS (France) driving simulator by testing an Extended Intelligent Cruise Control (EICC) both in a simulator and on a test track. EICC is intended to provide the driver with information about the headway between his car and the

leading vehicle. Data with and without the EICC were compared on both conditions and the three criteria used for the analysis were: the time headway, the heart rate and the steering reversal rate. The results showed that, in the simulator, it was more difficult to estimate distances accurately, particularly long ones (probably due to the poorer visual cues in the simulator or due to the lack of motion drivers could not estimate speed properly). Generally, steering, adjusting speed and estimating distances were very difficult in the simulator (especially the steering) whereas using the visual display was difficult in real life. He concluded that maybe a moving base simulator is required when speed adjustment are of primary importance in the task considered. He used a different questionnaire to test the sub-tasks. The results are presented in the following Table 5-4.

Sub-tasks	Instrumented vehicle on a test track	Driving simulator
steering	easy	very difficult
adjusting speed	easy	difficult
estimating distances	easy	difficult
coping with traffic	difficult	no traffic
bearing accelerations	high lat & longit acceleration	no acceleration
using the visual display	rather difficult	easy

Table 5-4 Relative difficulty of the different sub-tasks for the two systems

5.4.4. The RENAULT driving simulator validation study

Boulanger and Chevennement (1995) followed the validation approach of the RENAULT group described in paragraph 4.1 studied driver behaviour in the simulator and on a test track using an instrument vehicle. For the real road experiment they used 166 subjects divided into three different groups and dedicated to drive three different cars of the same model (Renault 19) but differently adapted to understeer or oversteer on an emergency manoeuvre test track. They wanted to test if results obtained from a fixed-base simulator can be compared with results obtained from this real road dynamic situation. The performance measures were speed, steering-wheel angle, yaw angle, longitudinal acceleration, lateral acceleration, brake pedal position, accelerator pedal and clutch pedal position. They also measured skin resistance; skin potential; skin temperature; heart frequency; breathing frequency and blood rate of all subjects. The independent variables were the two systems, the three different driving speeds (60, 70 and 80 km/h), the three different cars (three groups) and the three times (three laps) the performed the experiment. There were separators along to track to guide the drivers along a specific path. During the last lap, the separators were changed manually into a new position before the arrival of the car creating a sudden narrow curve and a "sudden throttle off". For the simulated experiment they used 66 drivers (22 drivers per each category of car). They also used a questionnaire about the subjective realism of the simulator, the one it is "always used during their different experimental campaigns". The results for the number of lane leavings showed that for each group at least 3 percent went out of the lane during the last lap on the right side and this percentage increased to 30 percent when it was an understeer or oversteer group. The understeer group got back into their lane quicker than the oversteer group. There were no lane leavings in the first two laps. Some of the findings are summarised in Table 5-5. Overall, for the simulated conditions, it was concluded that the understeer car is less suitable for sharp manoeuvres and that a fixed-base simulator cannot be so accurate in cases of small radius curvature, high visual yaw and high dynamic car behaviour.

Performance measures	Simulator			Problems
	standard	oversteer	understeer	
Speed variance	Speed ± 30%			longitudinal control
Steering-wheel operation variance			↑	lateral control
Steering-wheel rate and angle variance		↑	↑↑	lateral control
Lane leavings (both sides)	45%	65%	65%	
Simulator sickness	1.5%	6.1%	0%	

Table 5-5 Results of the simulated experiment

5.4.5. The TRL driving simulator validation study

Duncan (1995) investigated the validity of the TRL driving simulator by comparing individual drivers' performance of the same driving task in the simulator and a test track using an instrumented vehicle. The primary driving tasks included speed estimation, speed choice and lateral position, headway choice and maintenance and braking where the secondary tasks were eye glance behaviour, headway, and braking. All tasks were performed twice in both environments and the test track was driven both clockwise and anticlockwise. He also used a questionnaire for the subjective opinions of the subjects for the realism of the simulator. The subjects rated the "realism" and "usability" of the steering wheel response as between "poor" and "moderate" and commented about the tendency to "overcorrect" steering deviations. "Staying in the centre of the lane" was rated as the most difficult task in the simulator compared to the track where "Braking-Longer stopping Distance" was ranked second in difficulty. "Vehicle motion during braking" was rated the worst aspect of simulated motion and "Braking response-realism" as the worst aspect of vehicle control simulation. The results from the comparison of the two environments are presented in Table 9-1. It was also found that subjects over 40 years old are driving 1.83 mile/h slower than the younger subjects and female subjects drive in average 2.14 mile/h slower than the male subjects. Duncan took into account the perceptual correspondence of the TRL simulator by conducting a speed estimation experiment in both environments. Subjects had to drive the circuit of the test track three times trying to maintain a speed of 45 mile/h: the first time with the speedometer obscured, the second time without and the third time with the speedometer obscured again. This block of three trials was then repeated by driving the circuit the opposite direction. The analysis -using paired t-test comparisons- showed that the initial estimation of the speed was on the low side and did not differ significantly for both environments. The post speedometer estimate of speed was 46.54 mile/h for the simulator and 44.77 mile/h for the track (significantly different). Also, the mean speed increased significantly for both environments, especially of the simulator (+2.08 mile/h). An important finding was that the between-subjects speed variance was three times greater in the simulator than on the track.

5.4.6. The JARI driving simulator validation study

Soma et al (1996) investigated the behavioural and physical validity of the JARI moving-base driving simulator. They conducted their lane change test (double lane change course ISO-TR3888) on a test track and in the simulator. They used accelerator displacement, steering wheel angle, vehicle velocity, yaw velocity and lateral acceleration as dependent variables and the motion system on and off at the two conditions (test track and simulator) as independent variables. The results showed that when the motion system is on subject's driving speed (76.9 km/h) is closer to the field speed (77.8 km/h) than without motion (71.8 km/h) (in accordance

with Alm, 1995 results). The correlation analysis showed significant relation in "field test vs. simulator with motion" but not in "field test vs. simulator without motion". Hogema (1992) has also used the same lane change test to investigate the effect of visual delay in a simulator and a compensation technique to overcome this delay.

5.4.7. The HYSIM driving simulator validation study

Alicandri et al (1986) determined the absolute and relative validity of HYSIM driving simulator comparing real road and simulated data and using as a secondary task sign detection and recognition distance. HYSIM is a fixed-base, night-time scenario simulator with computer generated imagery for roadway definition and a wide screen TV projector and slide projectors to display the roadway. As dependent variables they used Average Sign Detection Distance, Standard Deviation of Detection Distance (the point at which a subject first reported seeing a sign of a specific background colour stated by the experimenter), Average Sign Recognition Distance (the distance form a sign at which the subject could read/understand the sign), Standard Deviation of Recognition Distance, Speed, Accelerator Position Changes (any change in the position of the pedal incorporating ≥10 percent of its total travel) and Steering Wheel Reversals (any movement of steering wheel exceeding 20 degrees). As independent variables they used age, sex, signs, the data collection zone and the two conditions.

	Av. DET. Distance (ft)			Av. RE. Distance (ft)		
	sim	**RR**	**F or t**	**sim**	**RR**	**F or t**
Place	498	659	F(1,27)=44.15, p=.0001	-	-	-
Sign by Place	1) 231* 2) 711 3) 639 4) 432	295 816 858 632	F(3,67)=4.88 p=.0041	1) 160 2) 568 3) 490 4) 359	193 501 443 433	F(3,68)=7.5 p=.0002
Variance Differences	1)61.49 2)60.67 3)65.56 4)39.33	57.40 246.44 244.86 175.46	 t=4.73, p=.0001 t=7.64, p=.0001 t=5.75, p=.0001	1)56.39 2)150.60 3)83.64 4)63.54	50.60 359.28 205.46 192.67	 t=2.52, p=.01 t=4.73, p=.0001 t=5.75, p=.0001
Sex (M, F) by Place	M=605, F=546		F(1,28)=3.15 p=.0869	M=393 F=395	M=441 F=344	F(1,28)=4.12 p=.052
Sex by Sign	-		-	longer for M than F for all signs		F(3,84)=2.59 p=.0571
Age	-		-	i) 434** ii) 367		F(1,28)=4.02 p=.0546
Differences between signs	1) 263 2) 763 3) 532 4) 748		F(3,68)=227.37 p=.0001	1) 176 2) 534 3) 466 4) 396		F(3,84)=105.35 p=.0001

*The numbers represent the four different type of signs that were used for the experiment:
1) Deer crossing, 2) Merge, 3) Parkway Headquarters and 4) School.
** The letters (i, ii) represent the two groups of ages: i) subjects under 30 years old and ii) subjects over 30 years old

Table 5-6 Results of the ANOVA for the average detection distance and the average recognition distance.

	Speed (mile/h)			APC		
	sim	**RR**	**F or t**	**sim**	**RR**	**F or t**
Place	no significant differences			0.462	2.49	F(1,28)=23.32 p=.0001
Sign by Place*	1) 35.0 2) 44.9 3) 44.5 4) 44.5 5) 45.6	32.6 45.8 44.7 40.1 50.6	F(4,80)=103.35 p=.0001	1) 0.22 2) 0.28 3) 0.78 4) 0.87 5) 0.16	2.64 2.53 3.16 1.97 2.13	F(4,105)=3.19 p=.0161
Zone by Place**	a) 41.4 b) 41.2 c) 41.2	43.0 42.3 42.9	F(2,40)=4.26 p=.0001			
Sex				APC greater in the field, ↑ APC for F in the field		F(4,105)=2.77 p=.0309
Differ. between signs				1) 0.22 2) 0.28 3) 0.78 4) 0.87 5) 0.16	2.64 2.53 3.16 1.97 2.13	F(4,112)=4.07 p=.0161

*The numbers represent the four different type of signs that were used for the experiment:
1) Deer crossing, 2) Merge, 3) Parkway Headquarters , 4) School and 5) Dummy zone.
** The letters a,b,c represent the three different zones were used for collecting speed and APC data: a) Zone A, b) Zone B, c) Zone C

Table 5-7 Results of the ANOVA for the average speed and the accelerator position changes

	SWR		
	sim	**RR**	**F or t**
Place	2.95	4.91	F(1,20)=79.57 p=.0001
Zone by Place*	a) ↓ b) ↓ c) ↓	↑ ↑ ↑	F(2,40)=53.93 p=.0001
Age**	group ii) more SWR	no dif. between groups i) & ii)	F(1,20)=5.41 p=.0306
Differ. between signs***	Average across sim & RR 1) 5.32 2) 3.05 3) 3.87 4) 4.62		F(4,80)=18.63 p=.0001
Differ. between zones	a) 4.35 b) 3.364 c) 3.80		F(2,40)=5.84 p=.006

* The letters a,b,c represent the three different zones were used for collecting speed and APC data: a) Zone A, b) Zone B, c) Zone C

** The letters (i, ii) represent the two groups of ages: i) subjects under 30 years old and ii) subjects over 30 years old
***The numbers represent the four different type of signs that were used for the experiment:
1) Deer crossing, 2) Merge, 3) Parkway Headquarters and 4) School.

Table 5-8 Results of the ANOVA for the number of SWRs per 1000 feet

Type of data	Correlations				
SIGNS	**DET. Dist.**	**RE. Dist.**	**Speed**	**APC**	**SWR**
Average data	r=.75 p<.0001	r=.53 p<.0001	r=.60 p<.0001	not signic.	r=.43 p<.0001
Group means	r=.96 p<.0374	r=.92 p<.0792	r=.86 p<.0001	r=.51 p<.0211	r=? p<.0002

Table 5-9 Results for the correlations relative to signs and the dependent variables

Conclusions

1. Signs with the greatest difference in variance between the HYSIM and the field show the greatest increase in the average detection distance in the field over the HYSIM. The smaller SD of the HYSIM scores is a result of the physical restrictions on the visual range inherent in the system and as expected indicated there is better control in the simulator (Table 5-6);
2. There were differences in speed between the HYSIM and the field;
3. Subjects' behaviour is different in the near vicinity of the sign. As they approach the sign the task load increase and hence, they slow down. After the sign was passed, and the detection/recognition task completed, the speed increased again;
4. There were more SWR per 1000 ft in the field than in the HYSIM;
5. The low number of "place" main effects in the ANOVAs suggests that there is some degree of absolute validity as well.

5.4.8. The UMTRI driving simulator validation study

Reed and Green (1995) investigated the validity of low cost driving simulator (UMTRI) by comparing driving on real road and in the simulator when dialling a phone number and using two visual scene fidelity levels in the simulator. The phone task was to dial three times an 11-digit long-distance number which was shown on a LCD screen mounted near the centre of the instrument panel. The low fidelity scene was black except the white road-edge lines and the centre dashed line. In the high fidelity was coloured and textured and there was also road environment. Due to technical problems, the data from using the phone on the low-fidelity scene were lost. Only 1 out of their 12 subjects complaint about simulator sickness.

The dependent measures they used for their analysis are shown in Table 9-1. Out of these variables, mean lateral speed and standard deviation of steering-wheel position were used as a measure of **lane keeping** and standard deviation of speed and standard deviation of throttle position as a measure of **speed control**. The comparisons presented in Table 9-1 are those for normal driving conditions only (not when dialling a number) and high fidelity visual scene. For the real road experiment, the analysis showed that for lane keeping, both task and age*task interaction were significant, but the gender related effects were significant and for speed control there were no effects of age or gender. When comparing the low and high fidelity visual scene, the analysis showed that the only significant difference was that the steering-wheel precision was opposite for men and women. Hence, they decided to compare the real road data only with the high fidelity simulated data. The results from this comparison showed that generally the arithmetic values taken from the simulator were greater than their counterparts from the real road, "indicating decreased driving precision in the simulator".

Specifically for **lane keeping**, the mean lateral speed was considerably higher in the simulator and age, task and age*task interactions were significant in both environments but the magnitudes of their effects were larger in the simulator. It was also found that the simulator is

more sensitive to the interfering effects of the secondary task, hence this "may make simulator studies more sensitive for detecting conditions of high driver workload than on-road studies". The age*gender and age*gender*fidelity interactions were significant. It was found that older subjects show a larger difference in driving precision between the car and the simulator during the phone task and older women show a larger difference than order men. For normal driving conditions the age*gender interaction was markedly smaller. For the standard deviation of the steering-wheel angle, the results were similar. In particular, older women showed significantly higher steering variance than other subject groups, particularly while performing the phone task. "Subjects were frequently observed to correct a lane-edge exceedance in the simulator with a large, rapid steering-wheel motion that would have produced tire squeal, high lateral acceleration and body roll, and possibly a loss of control on the road.... in a manner similar to the loss of control of an actual vehicle in slippery road conditions". For **speed control**, the effects of age, fidelity and task were different from the effects of lane keeping but the age effects were more important than the car/simulator differences. The variance of the throttle position was larger on the road than in the simulator.

The results from the correlation showed that "there was a much larger range of values across subjects in the lane-keeping variables, particularly because age had more pronounce effects on lane-keeping than speed control". According to Reed and Green their simulator "demonstrated good absolute validity for measures of speed control and good relative validity for the effects of the phone task and age on driving precision". Their results about older drivers, i.e. decrease on their driving performance when using a phone in the car are in accordance with other studies (Ponds et al, 1988). This means that "differences in multi-task performance ability between subject groups that would be significant on-road would produce even larger effects in the simulator".

5.4.9. The Daimler-Benz validation study

Riemersma et al (1990) investigated the validity of the Daimler-Benz driving simulator in evaluating speed-reducing measures. They used German drivers for the simulated experiment and genuine Dutch road users for the real road experiment. The speed-reducing measure (a traffic calming scheme) for the entrance of a Dutch village included two traffic signs (overtaking is prohibited and speed limit) 300m before the entrance of the village, a median strip 128m before the entrance of the village, a portal gate of yellow-coloured poles built by the entrance of the village and a different colour asphalt after the entrance of the village where the markings along the side of the road were removed. Two instructions were given when driving in the simulator: a) drive in a relaxed and unhurried manner and b) drive as quickly as the conditions would allow (under time pressure). The subjects drove 12 times the approach zone in the simulator for each driving instruction. As independent variables were used the longitudinal position of the vehicle on the road, the speed of the vehicle, the position of the accelerator as a percentage of total deflection and the braking force as a fraction of 400N. These data were recorded at a sampling frequency of 5 Hz.

An analysis of variance was performed to test the interactions between the approaching speed, the instruction and the different speed-reducing measures ("configuration"). The results showed that "configuration" had the largest effect ($F(1,20)=57.1$, $p<.000$, $r=0.28$), "instruction" had also a marked effect ($F(1,20)=23$, $p<.001$, r-15.6) and a slight interaction was found between the "configuration" and "instruction" ($F(1,20)=6.35$, $p<.02$). The results from the comparison of the two environments before and after the implementation of the traffic calming measures with regard to speed only are presented in Table 9-1. It can be seen that the speed reduction in the simulator was larger than that observed in real life, there is a

larger variation in speed in the simulator compared to real road speed and a larger variation in speed in the simulator between the different runs.

5.4.10.Perceptual validity of driving simulators

Staplin (1995) investigated the capability of young and older drivers to judge the last safe moment to initiate a left turn at an intersection of oncoming traffic, both in the field and with different types of visual display techniques. His experiment deals more with the perceptual and cognitive validity of a driving simulator with regard to these different techniques rather than with behavioural validity. He used three different types of visual display: a) a cinematic presentation (3000 lines) b) a video projection and c) TV monitor presentation (400 lines). All different visual displays were produced from a filmed (30 frames/s) approach of a white Mercury Marquis Sedan on a 2-lane highway at a speed of 48 km/h, from a perspective of a driver waiting to turn left onto an intersecting roadway.

The video images were National Television Standards Committee (NTSC) quality which theoretically permits 525 horizontal lines of resolution. In practice, production effects and transfer to laserdisc resulted in an effective resolution of less than 400 lines. The 35mm cinematic images, by comparison, displayed an equivalent horizontal resolution of over 3000 lines for 72 degrees horizontal field of view. The large screen display formats preserved correct size and perspective cues, such that the angular change associated with the target's motion in depth was consistent provided the same cues available to a driver viewing the scene through the windshield. The 20-in television monitor display compresses the target stimulus, however, and did not present absolute changes in angular size of the target that were accurate for its motion in depth as viewed under real-world conditions. Thus, the TV monitor trials presented relatively lower resolution images, without correct size and perspective information; the projection video trials presented correct size and perspective information, also at lower resolution; and the cinematic trials presented correct size and perspective information, at extremely high resolution.

For the video projection and cinematic presentation, a Fiat 128 body and frame was used, whereas for the TV monitor trials a single-seat driving buck consisting of a frame without external model was used. Subjects were asked a) to identify the earliest moment they can see the target car by pressing a button on the steering wheel and b) to depress the pedal at the last possible safe moment to turn in front of the target vehicle in the simulated experiment. For the real road experiment a hand-held response button was used to identify the when the target vehicle reached the last possible safe moment. The measurement system was accurate to the nearest foot. Due to uncontrollable real road traffic, the experimenter was driving the car and the subject was sitting on the passenger seat. When he positioned the car properly at the intersection, he radio-contacted the target vehicle to start the approach. Data were collected only when no other traffic was present but to avoid any confusion the experimenter made known to the subject when the target vehicle was approaching therefore no target recognition distance data were obtained from the real road experiment unlike the simulated experiment.

The results from the comparison between the three different visual displays in the simulator and the real road showed that for the cinematic presentation the yielded results were similar to those in the field test and in particular, an increase in the judged minimum safe gap with increasing target approach was obtained ($F=14.28$, $df=1$, $p<.0009$). The comparison of the two other displays and the field test yielded different results. The general finding was that older people were relatively insensitive to the speed of approaching vehicles. Staplin concluded that "tasks that critically depend on estimation of speeds and time duration may be

concluded that "tasks that critically depend on estimation of speeds and time duration may be affected by image resolution limitations". These findings are in line with the results of Kaptein et al (1996) that, in a typical mid-level driving simulator, judging when to start braking was relatively difficult at large distances (even modern driving simulators do not present image resolutions that are achieved with cinematic presentations. Generally, the resolutions which are used are similar to those for video projection). Staplin also suggested that "image/scene attributes including high resolution and correct size and perspective cues may be prerequisites for valid and generalizable driving simulation measures of visual sensory/perceptual task performance". The results from the real road and simulated experiment are presented in the following tables. The following abbreviations are used: sample size (n); mean (M); standard deviation (SD).

Real Road Experiment						
"Least Safe Gap" Distance (ft)						
	Instrumented vehicle					
Age	Target Speed 30			Target Speed 60		
	n	*M*	*SD*	*n*	*M*	*SD*
1 (33.3y)	12	327	162	11	433	171
2 (65.1y)	13	512	164	11	519	108
3 (79.4)	14	546	203	15	527	173

Table 5-10 Mean and standard deviation of "Least Safe Gap" distance for the real road experiment for different age groups and number of subjects

Simulated Experiment																		
Target Recognition Distance (ft)																		
	TV monitor display						Projection video display						Cinematic display					
Age	Target Speed 30			Target Speed 60			Target Speed 30			Target Speed 60			Target Speed 30			Target Speed 60		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
1	22	864	138	22	750	154	25	971	150	25	814	182	23	1109	57	23	1120	91
2	26	890	166	26	762	163	28	975	187	28	855	224	25	1050	120	24	1101	67
3	21	872	111	21	713	186	24	999	150	23	850	184	20	1018	105	21	1046	92
"Least Safe Gap" Distance (ft)																		
1	22	312	112	22	327	95	25	480	182	25	487	152	23	536	169	23	661	165
2	26	451	184	26	406	139	28	748	220	28	630	223	25	666	221	25	753	184
3	21	448	182	21	389	130	24	830	156	24	728	189	21	708	235	21	750	202

Table 5-11 Mean and standard deviation of target recognition distance and "Least Safe Gap" distance for the simulated experiment for different age groups and number of subjects

Source: Table 5-10 and Table 5-11 are adapted from Table 1 of Staplin (1995).

The comparison of the different visual displays in terms of resolution, correct size and perspective cues showed that the cinematic display (which provides the best resolution and perspective cues) is the best because target recognition distance did not decrease for any age group for the lower versus the higher target approach speed. It was observed that older people are more sensitive to loss of spatial information cues (able to identify the real size of the target and the distance between them and the target) because they scored worse in judging the safe gap acceptance when using the TV monitor and the video projection displays.

Comparative results of all the above behavioural validation studies are summarised in Table 9-1, APPENDIX A.

6. COMPARISON OF THE EARLY AND RECENT VALIDATION STUDIES

The objectives of this comparison is to test, in a qualitative way, the hypothesis that a) high-cost simulators are better than medium-cost simulators and the latter better than the low-cost (part-task) simulators and b) moving-base simulators are better than fixed-base with hydraulic actuators simulators and the latter better than the fixed-base simulators. Before moving to the comparison of the behavioural validation studies to test if these objectives are true, we should first define the different types of driving simulators mentioned above and take into serious consideration the various assumptions underlying these studies and generally any validation study conducted on a driving simulator.

The definitions of low, medium and high cost driving simulators are given below. A similar classification is low-level, mid-level and high-level driving simulators (Weir and Clark, 1995).

1. Low-cost driving simulators can provide reasonable fidelity in the visual, auditory and control feel cueing. They called low-cost due to the relatively inexpensive graphics displays. They have the ability to move back and forth the simulator software from the desktop to the laboratory environment and they are particularly cost effective for students and dissertation related projects and vehicle manufactures and parts suppliers who are looking to support research on limited budgets.

2. Medium-cost driving simulators employ advanced imaging techniques (using real-time animation to create a scene that is projected in front of the driver), a large projection screen, a full-sized and complete vehicle with all the normal controls. Low and medium cost driving simulators can be either fixed-base (no kinaesthetic feedback) or can provide trivial motion feeling by using systems which simulate the normal vibrations experienced while driving and provide minimal car cab pitch for each corner of the car cab.

3. High-cost driving simulators provide an almost 360 degrees field of view and an extensive moving base. The motion system may include more than six degrees of freedom hexapod and it is built using the aircraft flight simulators technology. The translational motion capability can be greater than 2m (Weir and Clark, 1995).

The various assumptions wrongly accepted as being true in different validation studies could be summarised as follow:

1. The physical validity of the driving simulator

   It is very often wrongly assumed that there is absolute physical correspondence between the driving simulator and the actual vehicle which its technical characteristics have been simulated. The physical validity of the driving simulator is a prerequisite that should never been disregarded or mistreated, otherwise we can never be sure about the behavioural validity of the simulator.

   Especially when using instrumented vehicles for the real road experiment, it is assumed that the physical characteristics of that vehicle are the same as the simulator's. It cannot be the case when e.g. a Metro Rover is used as the instrumented vehicle and a Rover 216 Gti as the simulated vehicle. The capabilities and the size of these two vehicles are obviously not the case.

2. Real road data collection methods

It is assumed that the data collected from the real road are free of errors. This is not exactly true. The accuracy of the methods collecting real road data which are later compared with simulated data has to be taken into consideration. The traditional traffic engineering road data collection methods had almost been the same from the time the first behavioural validation studies started and there has been small improvement in the accuracy with which the data are measured until today. The use of instrumented vehicles for real road data collection seems to increase the accuracy of the data and make the comparison with the simulated data easier.

2. Differences between the real and the simulated environment

It is assumed that the simulated and the real road environment have been built as close as it can be to the real road. This cannot not always be true because it depends on various elements which are not always predictable, measurable and easy to define all of their parameters, such as:

a) type of the driving simulator (e.g. moving-base)

b) method for building the simulated road

- using specifically built-in house graphics software and

i) data obtained from an instrumented vehicle

ii) data obtained from traditional traffic engineering methods (e.g. road tubes, video cameras)

- using of the shelf software (e.g. MultiGen)

c) method for building the objects (video from real road and built-in house software; of the shelf software, e.g. MultiGen)

d) ability of the simulator to simulate real road environment

It is assumed that the "other" real road traffic has been simulated as close as it could be in the simulator. This depends on:

i) simulated traffic modelling

ii) drone traffic and event traffic

iii) software used for modelling the traffic

iv) validity of the software used to simulate the traffic

3. Number and homogeneity of subjects

Although this aspect may not seem so relevant, it is very important and it should not be neglected. It is assumed that using e.g. only males as subjects to drive the simulator, any results taken can apply generally to the driving population. It is obvious that this cannot be true. It cannot also be true that the results obtained from simulator experiments using very limited number of subjects can be valid for the whole driving population. The variability of these subjects is too great.

Although the above assumptions do not cover every assumption taken in any behavioural validation study, it is apparent that these assumptions have to be precisely stated in the beginning of any study and it would be better before starting the behavioural study to have at least the closest physical correspondence of the simulator to the actual car.

having the above on mind, the comparative results of the early behavioural validation studies regarding the absolute and relative correspondence between the real road and the simulated environment showed that although absolute correspondence was very poor, relative correspondence was high, i.e. the same trends in driving performance was observed in both

environments. These findings were based on five validation studies using driving simulators with limited computing and image generating subsystems and also low face validity.

The comparative results of the recent behavioural validation studies regarding the absolute and relative correspondence between the real and the simulated environment showed also that absolute correspondence was poor and relative correspondence was high. The difference is that these findings were based on twelve validation studies but the crucial element here is that this time the driving simulators used had advanced computing and image generating subsystems and also higher face validity than their counterparts in the early validation studies. From this rather simplistic comparison between early and recent validation studies regarding the absolute and relative correspondence of the two environments, one could conclude that despite the great developments and achievements of technology, not much more has been achieved in the driving simulators regarding their behavioural validity and the results relative to driving performance and that the abilities of the old scale model, part-task driving simulators could compete satisfactory with the new fancy driving simulators.

The author considered necessary to seek further this ambiguous finding by comparing the recent behavioural validation studies and to prove or disprove the hypothesis that medium-cost simulators are better than the part-task or low-cost counterparts and high-cost are better than medium-cost driving simulators.

6.1. COMPARISON OF THE RECENT VALIDATION STUDIES ACCORDING TO DRIVER BEHAVIOUR AND VALIDATION CRITERIA

In this paragraph the recent behavioural validation studies will be compared relative to the previously described driver behaviour levels (see paragraph 5.2) and validation criteria (see paragraph 4.3). The objective of this comparison is to test if there is any significant difference, in a qualitative manner, between the three different cost-type driving simulators. The recent behavioural validation studies will be compared relative to the type of simulator (fixed-base, fixed-base with limited motion and moving-base); the validation criteria (absolute and relative); the driver behaviour levels (control, tactical and strategic as they defined in paragraph 5.2) (DBL); the "type of driving" used in each validation study (as it is defined in the following paragraph by the author) (ToD); and the use of subjective criteria about the realism of the specific driving simulator (USB). The results of this comparison are given in Table 6-2.

The following table gives a short-coding for each of the "type of driving" used in each behavioural validation study (A, B, C, D). For example, if only braking was measured then is type A, if braking on dry versus icy road conditions was measured then it is type C, if braking was measured according to a rule e.g. if lead vehicle brakes to less than 1 second then break, it is type B and finally if the like in type B there was a distinction in the measurements between young and old drivers, then it is type D.

	Driver beh. level	**rule**
absolute	**A** (e.g. braking)	**B** (e.g. braking if + the rule: TTC)
relative	**C** (e.g. braking on dry v. icy conditions)	**D** (braking if + the rule + young v. old drivers)

Table 6-1 "Type of driving" used in the validation studies

Validation Studies	Driving simulator (type)	Validation criteria	DBL	ToD	USC
Wheaton et al (1966)	part-task	low absolute high relative,	tactical	A	yes
Wojcik and Weir (1970)	scale model	relative	tactical	A	no
Breda et al (1972)	?	poor absolute, good relative	strategic	B	no
Allen and O'Hanlon (1979)	TV monitor, FB	relative	tactical	A	no
Blaauw (1982)	TNO (scale model), FB, MC	relative for LP and absolute for S	control	D	yes
Kappé and Körteling (1995)	TNO (updated version), FB, MC	relative for LP and absolute for S	control	B	no
Tenkink (1989; 1990)	TNO (updated version), FB, MC	relative for S and SD_LP	control	B	no
Tenkink and Van der Horst (1990)	TNO (updated version), FB, MC	relative for S	control	A	no
Janssen et al (1991; 1992a,b) and van der Mede and van Berkum (1993)	TNO (updated version), FB, MC	absolute	strategic	A	no
Hogema (1992)	TNO (updated version), FB, MC	relative for ratings and SRR	tactical	B	
van der Horst and Hoekstra (1993) and Hogema et al (1993)	TNO (updated version), FB, MC	relative	N/A*	N/A*	no
Kaptein et al (1996) and van der Horst (1990)	TNO (updated version), FB, MC	absolute for low S and hard braking; relative for high S and normal braking	tactical	D	no
Alicandri, Roberts, Walker (1986)	HYSIM, FB, night time, MC	relative	tactical	B	no
Riemersma et al (1990)	Daimler-Benz, MB, HC	relative for S	control	B	no
Harms (1993)	VTI, MB, HC	relative	control	A	no
Alm (1995)	VTI, MB, HC	relative	control	A	yes
Harms et al (1996)	VTI, MB, HC	relative	control	B	no
Duncan (1995)	TRL, FB_H, MC	high face validity	tactical	D	yes
Malaterre (1995)	INRETS, FB, MC	relative	tactical	C	yes
Reed and Green (1995)	UMTRI, FB, LC	absolute for S and relative for	tactical	D	no
Boulanger, Chevennement (1995)	Renault, FB, MC	relative	tactical	D	yes

*this was a perceptual validation study

Table 6-2 Comparison of validation approaches, methodologies and criteria between the different validation studies

It can be seen that researchers usually employ the tactical level as first choice and the control level as second choice and very rare the strategic level when they want to investigate driving performance in the simulator and on real life. The use of questionnaires of the subjective realism of the simulator and the mental workload is not a common practice and whether the

simulator is low, medium or high cost the results always show relative validity for the variables that have been chosen to test the behavioural validity of the simulator.

6.2. COMPARISON OF THE RECENT VALIDATION STUDIES RELATIVE TO DRIVING PERFORMANCE

The recent behavioural validation studies which have been described in paragraph 5.4 will be compared in terms of driving performance (see paragraph 4 for definitions), i.e. the most commonly dependent variables and types of statistical analysis used in these studies.

Jovanis (1995) in his literature review also identified the most commonly dependent variables and statistical procedures employed for simulator studies. These variables included car following headway, lateral position, accuracy (error data or error frequency), reaction time, eye movements, vehicle speed, accidents and mental workload measures. ANOVA and/or regression analysis were mainly used for the statistical analysis. He argued that "new and more imaginative experiments must be conducted using more advanced statistical analysis methods" and that "the statistical techniques should allow for multiple dependent measures considered jointly not separately".

The following tables (Table 6-3, Table 6-5, Table 6-6, Table 6-7 and Table 6-8) compare the twelve behavioural driving simulator validation studies which have been investigated here (see Table 9-1, the study of Staplin (1995) is not included in the comparison) with regard to the number of subjects, the use of training sessions or not, the type of statistical analysis used, the three most commonly used dependent variables and the three most commonly used independent variables. Six of them conducted on fixed-based simulators, five on moving-base and one in a fixed-based with hydraulic actuators driving simulator.

Variable	min	mean	max
No of subjects	7	20	48

Table 6-3 The min, mean and max number of subjects used in the twelve validation studies

Variable	Real road and genuine road users	Real road and Instrumented vehicle	Test track and instrumented vehicle
Real road experiment	1 (8.3%)	6 (50.0%)	5 (41.7%)

Table 6-4 Type of real road experiment

Variable	yes	no	N/A
Training sessions	8 (66.7%)	1 (8.3%)	3 (25.0%)

Table 6-5 The use of training sessions in the twelve validation studies

Variable	ANOVA	CoM	Correlations	Other tests
Statistical analysis	9 (75.0%)	8 (66.7%)	6 (50.0%)	3 (25.0%)

Table 6-6 The type of statistical analysis used in the twelve validation studies

Variable	Speed	Lateral position	Steering behaviour*
Dependent variables	9 (75.0%)	7 (58.3%)	6 (50.0%)

* Steering behaviour means either steering-wheel angle or steering-wheel reversal rate

Table 6-7 The three most commonly used dependent variables in the twelve validation studies

Variable	Two conditions	Driving instructions	Moving system on-off
Independent variables	12 (100.0%)	5 (41.7%)	2 (16.7 %)

Table 6-8 The three most commonly used independent variables in the twelve validation studies

From the above tables it can be seen that on average twenty (20) subjects are used for either the simulated and/or the field trial and the majority of the validation studies has been conducted using an instrumented vehicle either on the real road or a test track. Only one study compared the simulated results with results obtained from genuine road users. The three most commonly used dependent variables are speed, lateral position and steering performance, the most commonly used type of statistical analysis is the analysis of variance (ANOVA) and besides the comparison of the two conditions (field and simulator trials), a number of researchers investigate different instructions in driving (e.g. slow v. fast) between the two conditions.

In the following paragraphs more emphasis will be given in the above findings and especially in their interpretation and applicability in real life driving conditions. The importance of the correct type of statistical analysis and the interpretation of the outcomes will be discussed too.

6.2.1. Statistical analysis

From Table 6-6 it can be seen that the most commonly used type of statistical analysis is the analysis of variance, then the comparison of means and then the correlations. In the following paragraphs, besides the above mentioned statistical analyses, the importance of null hypothesis will be introduced.

6.2.1.1.The Null Hypothesis

One critical issue in Human Engineering research is the use and interpretation of the null hypothesis (Ellis, 1967). The null hypothesis depends on the dependent variables, i.e. the performance measures which are observable and recordable representations of task relationships, underlying the man/machine interplay being studied (common sense and knowledge of previous research are important here) and the apparatus used for measuring task performance during experimentation must be sensitive to small but meaningful changes in the independent variables.

Crucial questions are:
1) What does it mean when Ho is accepted as tenable?

From a statistical standpoint this simply means that Ho, based upon the obtained experimental differences, is not the only tenable hypothesis, i.e., no one of a whole series of differences from less than obtained differences to near zero is also tenable. Depending upon one's experimental objective, this may have serious implications for data interpretation. For example, if the experimental objective is to determine absolute minimum differences between two study conditions, one must be careful in interpreting data results as representing these minimums on the basis of accepting Ho. However, if determining whether or not difference do actually exist between two study conditions is the experimental objective, then accepting or rejecting Ho is relevant evidence.

2) What if a false null is accepted (i.e. the statistical test being used accepts Ho, but the magnitudes of the differences are quite large?)
The answer to this question has direct bearing upon the relationship between what might be called the practical and the statistical significance of the data. For example, a high power statistical test in conjunction with a low a level and a large sample size could indicate that a difference of 0.001 rad/sec in pitch rate is statistically significant, but a guidance and control engineer will strongly argue that an error of this magnitude is certainly not crucial in the control of pitch for normal manoeuvres. In contrast, a low power statistical test in conjunction with a high a level and a small sample size may indicate that a pitch rate error of 0.1. rad/sec is statistically insignificant; however, from a practical standpoint an experienced test pilot or engineer knows that errors of this magnitude are crucial.

Therefore, to properly interpret Ho based on obtained data, one must have previously given attention in experimental planning phases to developing and maintaining a high correlation between statistical and practical significance. Suggested recommendations for accomplishing this objective from a statistical standpoint include the following:

1) Select a statistical test on the basis of its power (most powerful test is preferred) and of its applicability to the underlying scale of measuring, i.e. nominal, ordinal or interval.
2) Use an a level equal to 0.05 in conjunction with a two-tailed test of significance. Smaller a levels can be deceiving when Ho is the experimental hypothesis. In fact strong arguments can be made for enlarging the meaningful range of a from 0.05 to one of a series of values as large as 0.10.
3) Use the smallest sample size (n) possible as determined by solving for n in the statistical test equation where a and the magnitude of difference one is willing to accept as having practical significance are known.

From the standpoint of practical significance of data, the overriding recommendation is simply this: Depend upon the expert knowledge of other technical disciplines.

Alicandri et al (1986) used the null hypothesis with the probability of making a Type II error to be minimal because it will be more costly to fail to detect a significant difference when a significant difference existed than to reject the null hypothesis, when in fact, the null hypothesis was true. To accommodate this requirement, an alpha level of 0.10 rather than the more conventional 0.05 level, was utilised in the analyses. The reason was to identify those aspects of system functioning which were not producing valid simulations. Subsequent system modifications could then be undertaken to bring system operation into closer correspondence with real-world conditions.

Reed and Green (1995) used statistical effect tests with Type-I error probabilities less than or equal to 0.01 are considered significant.

6.2.1.2.Analysis of Variance

Alicandri et al (1986) used a 5-way analysis of variance to interpret their data. They used PROC GLM (General Linear Model) where there were missing data. If the distribution of the missing data was such that if two subjects are randomly dropped, the cell sizes would be equal at n=6 and given the statistical parsimony and computational economy of a balanced-cell design, the subjects were dropped and PROC ANOVA of SAS was used. On the assumptions when using Analysis of Variance is that the variances of the various sub-groups be equal and, in the case of factors using repeated measures, that the co-variance matrices be equal. The Greenhouse-Geisser correction (Winer, 1962) was used as a more conservative test in case where the usual F-test was fairly low. This conservative test is mentioned only when it resulted in a probability level of greater than 0.10 and the usual test resulted in a level less than 0.10. Those result should be regarded as having borderline significance.

6.2.1.3.Correlations

Alicandri et al (1986) measured two types of correlations a) by using averaged data and group means and b) fined-grained correlations. Low values for fine-grained correlations do not necessarily mean that the simulator is not valid. On the other hand, in combination with good correlational values from averaged data, they show that the simulator is valid as long as certain adjustments are carefully planned in the experimental design to preclude restricting the range of conditions under which data are collected. Correlations were run on the complete set of data using SAS's PROC CORR NOMISS. With this option, if one of a pair of data points is missing, the other is dropped from the analysis.

Blaauw (1982) used the Pearson product moment correlations. The small correlations he found did not necessarily indicate differences between the dependent variable and the two environments because as he stated "the homogeneity of the groups of subjects with respect to the specific variables could produce a restriction of range".

6.2.2. Driving speed

Higher speeds in the simulator compared to real life have been observed either on straight (Blaauw, 1982) or curved road sections (Tenkink, 1990; Tenkink and Van der Horst, 1990; Harms, 1993; Duncan, 1995; Harms et al, 1996).

It should be noted here that Harms et al (1996) observed higher speeds in the simulator on a road of smaller width and having obstacles (tunnel walls) positioned exactly by the road edge compared to the previous VTI studies (larger road width, no nearside obstacles) This is contradictory with Tenkink (1990) and Tenkink and Van der Horst (1990) findings, i.e. they observed that in both systems driving speed reduced when obstacles were placed nearer the road and with decreasing road width and decreasing curve radius.

Riemersma et al (1990) found that the speed reduction in the simulator was larger than that observed in real life after the implementation of speed reducing measures. Alicandri et al (1986) also did not find any significant differences in speed between the HYSIM and the field, although subjects drove slightly faster on the road. Green and Reed (1995) found that the variance of the throttle position was larger on the road than in the simulator, meaning that it was easier for the subjects to keep a steady speed in the simulator.

There are numerous explanations why there are differences between the real and the simulated environment. The author will present here a summary of various reasons proposed by the researches who have conducted the behavioural validation studies:

1. Differences between the simulated and the real road geometry/environment/layout/other road users (Alicandri et al, 1986; Tenkink, 1990; Riemersma et al, 1990; Green and Reed, 1995);
2. Differences between the face validity (size, capabilities, engine noise) of the instrumented vehicle and the simulated vehicle (Alicandri et al, 1986; Tenkink, 1990; Green and Reed, 1995);
3. Lack of acceleration forces for the fixed-base driving simulators (Tenkink, 1990);
4. Lack of visual information in the simulator (Tenkink, 1990);
5. The different type of subjects used on real road and in simulator experiment (they are used to different speed limits) (Riemersma et al, 1990);
6. The different time periods for the after period of comparison between the simulator and the real road (Riemersma et al, 1990);
7. The different instructions were given for the simulated driving (Riemersma et al, 1990);

The different type of speedometers (analogue gauge in the car and digital LED panel in the simulator) (Green and Reed, 1995).

Speed variation

Speed variation (or variance) has been considered from time to time more important than driving speed for traffic safety from a number of researchers (including the author of this paper). The reason is the findings of Solomon (1964), Cirillo (1968) and Hauer (1971) and Blana (1994), that increase of speed variance leads to increase of traffic accidents. What is usually observed in driving simulators is that differences in speed between the real road and simulated environment are not statistically significant or there are minor differences whereas variation is speed is always statistically different between the two environments.

Increase in speed variation in the simulator has been reported by Riemersma et al (1990), Harms (1993), Alm (1995), Duncan (1995), Reed and Green, (1995), Boulanger and Chevennement (1995), and Harms et al (1996).

Speed estimation

Duncan (1995) found that there were no significant differences between the two environments for the initial estimation of the speed but the post speedometer estimate of speed was significantly higher in the simulator compared to the track.

6.2.3. Lateral position

Green and Reed (1995) found that "there was a much larger range of values across subjects in the lane-keeping variables, particularly because age had more pronounce effects on lane-keeping than speed control". Harms (1993) and Blaauw (1982) have observed statistically significant differences on lateral position between the two environments. Harms (1993) initially suggested that this problem could be due to the absence of other traffic, or that the subjects use other visual cues for their lateral control in a driving simulator than during field driving. Alm (1995) did not find any statistically significant differences on average speed and lateral position (there was opposing traffic this time) for both environments but differences were found in the last VTI behavioural validation study (Harms et al, 1996).

Lateral position variation

Increase in lateral position variation in the simulator compared to driving on real road has been reported by McRuer and Kendal (1974), McLane and Wierwille (1975), McRuer and Klein (1976), McRuer et al (1977), Blaauw (1982), Tenkink (1990), Harms (1993), Alm (1995) and Harms et al (1996).

In particular, Blaauw (1982) in his validation study, trying to separate the effects of additional time delay and the absence of kinaesthetic feedback, found that even though time delay has been minimised, the problem of the increased standard deviation of lateral position was still there. He concluded that "drivers performed more poorly in the fixed-base simulator due to a diminished perception of lateral translation (absence of kinaesthetic information)". Tenkink (1990) found that in both systems variation in lateral position decreased when obstacles were placed nearer the road, although higher in the simulator.

Harms et al (1996) concluded that "the presence of critical but unnoticed source of variance, influencing subjects speed and lateral position both in the field trials and simulator trials, may result in unreliable conclusion of behavioural validation studies".

6.2.4. Steering behaviour

The steering behaviour is importance for traffic safety because it is related to the control of the vehicle, for example high steering reversal rate means high driving task demand.

Blaauw (1982) found that subjects steer at higher frequencies and in a more oscillatory fashion in the simulator than in the instrumented vehicle on straight road sections.

Reed and Green (1995) found that lane keeping (mean lateral speed and standard deviation of the steering-wheel angle) was considerably higher in the simulator, i.e. subjects drove with greater precision in real life than in the simulator, which is in accordance with other studies (McRuer and Klein, 1975; Blaauw, 1982; Hogema, 1992; Harms, 1993). They observed that subjects were frequently correcting "a lane-edge exceedance in the simulator with a large, rapid steering-wheel motion that would have produced tire squeal, high lateral acceleration and body roll, and possibly a loss of control on the road... in a manner similar to the loss of control of an actual vehicle in slippery road conditions". On the other hand, Alicandri et al (1986) observed more SWR per 1000 ft in the field than in the simulator.

Boulanger and Chevennement (1995) concluded that for the simulated conditions, the understeer car was less suitable for sharp manoeuvres and that a fixed-base simulator cannot be so accurate in cases of small radius curvature, high visual yaw and high dynamic car behaviour.

Possible reasons for the observed differences of steering behaviour between the two environments could be (either positive or negative):

1. differences in handling and dynamic characteristics of the vehicles used for the field and the simulated experiment) (Alicandri et al, 1986);
2. the lack of other traffic (other vehicles may have caused subjects to perform more SWRs in the field) (Alicandri et al, 1986);
3. differences between the real and the simulated environment (e.g. winds and road humps texture which may have caused the increase to the no of SWRs in the field could not be simulated) (Alicandri et al, 1986);

4. the lack of vestibular cues and the absence of danger (Reed and Green, 1995);

These problems could be overcome by:
1. "Reducing the yaw inertia of the simulated vehicle might help to reduce the tendency of the subjects to overcorrect their yaw errors" (Reed and Green, 1995);
2. "... the steering system itself might be programmed to have substantial resistance to large-magnitude, high-frequency movements, reducing the possibility that the subjects will begin an unstable correction plan" (Reed and Green, 1995).

6.2.5. Braking performance and headway

Duncan (1995) observed, during braking performance, a tendency to over-compensate for the lack of real motion by excessive use of controls.

Duncan (1995), Malaterre (1995), Staplin (1995), and Kaptein et al (1996) found that braking over a long distance is one of the most difficult aspects of driving in the simulator. Possible reasons can be limited image resolution and the lack of kinaesthetic information (Staplin, 1995; Kaptein et al, 1996).

Staplin (1995) observed an increase in the judged minimum safe gap with increasing target approach was obtained where Duncan (1995) found that fixed and safe headway were regarded as particularly difficult in the simulator.

6.2.6. Learning effects and sequence of environments

Learning effects

Learning effects in the simulator appeared to be similar to those associated with the instrumented vehicle: seconds attempts at a task were driven slightly faster, although steering quality did not significantly improve. When braking to a target which was a control task not covered in the familiarisation, subjects in the simulator did exhibit a rapid learning effect as they adjusted to the different balance of cues available (Duncan, 1995).

Soma et al (1996) found that in the field and the simulator with motion, the ration of constant speed increases when increasing the trial number but not in the simulator without motion. This means that subjects increase their ability to control the moving simulator and the instrumented vehicle with increase number of sessions but not in the static simulator.

Sequence of environments

Blaauw (1982) found that the standard deviation of lateral position and mean driving speed increased significantly when subjects transferred from the first environment to the second environment during the day (in.ve.(*16.7cm*)→ sim(31.9cm) v. sim(28.8cm)→in.ve.(*19.3cm*); in.ve.(*104.1 km/h*)→sim(109.8 km/h) v. sim(101.2 km/h)→ in.ve.(*107.0 km/h*) respectively).

6.2.7. Scene complexity and road environment

Kaptein et al (1996) compared low and high fidelity visual scene and they found that "scene complexity showed not to be important" but "field of view is important during the braking manoeuvre". In particular, the results showed that with a simple scene the stopping distance

decreased with field of view whereas with complex scene the stopping distance increases with field of view. No explanation could be given from the authors for this finding.

Reed and Green (1995) compared low and high fidelity visual scene (frame rate 20Hz-25Hz v. 40 Hz) and the analysis showed that the only significant difference was that the steering-wheel precision was opposite for men and women.

Staplin (1995) compared three different visual displays (TV monitor, video projection and cinematic displays) in terms of resolution, correct size and perspective cues. The results showed that the cinematic display is the best because it yielded the best results compared to the field trials.

6.2.8. Moving system

Alm (1995) found significant difference in lateral position between the moving base off and the other two conditions (real road and moving system on) but no effect in speed or the variation in speed level whether the moving system was on or off. On the other hand, Soma et al (1996) found that when the motion system is on, subject's driving speed is closer to the field speed than without motion.

Alm (1995) concluded that the moving-base system minimises the nausea effects from the simulated road environment and helps the driver to keep the car on a steady course on the road.

6.2.9. Mental workload

It is not exactly known if driving in a simulator –with or without a secondary task– increases the mental workload of drivers and therefore their ability to control the car with accuracy. Researchers usually use the NASA-TLX or built-in-house questionnaires to check this aspect.

Alm (1995) found that driving in a simulator was more physically demanding, more effort demanding, and more frustrating than driving on a similar real driving condition, therefore it produces higher mental workload compared to real car driving.

Duncan (1995) and Reed and Green (1995) found that driving with a secondary task was more demanding in the simulator than on the real track, causing the subject's performance to be more sensitive to additional tasks. This suggests that the subject on the track had reserves of attention and capability which could be called upon to refine steering behaviour or deal with distractions. In the simulator, it appears that the subject was working close the limit of their resources and abilities, so that improvements in performance were more difficult to achieve and limited attention had to be split between task, resulting in increased primary task degradation (Duncan, 1995). Simulator studies are more sensitive for detecting conditions of high driver workload than on-road studies (Reed and Green, 1995).

On the other hand, Alicandri et al (1986) found that the task load remains high in the field, but decreases in the simulator after the primary task (detection/recognition of the sign) had been performed.

6.2.10.Realism of the simulator

The realism of the simulator is usually checked using questionnaires of the impressions and opinions of the subjects driving the simulator. The researchers use different questionnaires and different scales for the ratings. The issue of realism is critical for the face validity of the simulator.

In Blaauw's (1982) questionnaire, all drivers rated the simulator more unfavourable (task difficulty, required attention and monotony) compared to the instrumented vehicle with the exception of the longitudinal control (driving on a straight road with no other traffic). The subjects also commented about the monotony of the simulator due of lack of other traffic, road curvature and road signing.

In Malaterre's (1995) questionnaire, subjects rated steering, adjusting speed and estimating distances (particularly long ones) as very difficult in the simulator (especially the steering) whereas using the visual display was difficult in real life. He concluded that, probably due to the poorer visual cues in the simulator or due to the lack of motion, maybe a moving base simulator is preferable when speed adjustment is of primary importance in the task considered.

In Duncan's (1995) questionnaire, subjects rated the "realism" and "usability" of the steering wheel response as between "poor" and "moderate" and commented about the tendency to "overcorrect" steering deviations. "Staying in the centre of the lane" was rated as the most difficult task in the simulator compared to the track where "Braking-Longer stopping Distance" was ranked second in difficulty. "Vehicle motion during braking" was rated the worst aspect of simulated motion and "Braking response-realism" as the worst aspect of vehicle control simulation.

Alm (1995) found that there were no significant differences between the realism of the simulator when the moving system was on or off, still all the ratings had the same tendency that the realism is better when the moving system is on. Driving on straight road sections with the moving system on or off wasn't very different, where driving on curves was rated more positive with the moving system on.

6.2.11.Interpretation of results

The design and interpretation of simulator experiments must take into account the increase variability of subject reactions in the simulator and the particular demands associated with simulating heavy braking and cornering forces in a fixed-base environment (Duncan, 1995).

7. CONCLUSIONS

Low-cost versus medium-cost and high-cost driving simulators.

Caro (1973) stated that "there is substantial applied research evidence that much of the training being conducted in expensive simulators could be accomplished in less expensive devices if the training programs used with them were properly designed and conducted". Evans (1991) concluded that "high realism simulators appear to offer little for driver training, although rudimentary low-cost simulators can be useful in initial instruction of location and function controls...While the performance skill learned in simulators can be critical in emergencies in the air, car driving emergency situations usually arise because of expectancy

violations". This literature review, although broadly related to research driving simulators and not training simulators showed that these statement can be true. Differences between the real and the simulated environment related to driving speed, lateral position, variation in speed and lateral position, steering behaviour and mental workload emerge whatever the cost of the driving simulator. Scene complexity does not seem to improve simulated driving performance where there is not enough evidence about the importance of field of view. It seems that the most important element for a successful behavioural validation study is the carefully designed experimental procedure, including the statistical analysis, and the correct interpretation of the results.

Fixed-based versus moving-base driving simulators

There is not enough evidence to support the hypothesis that moving-base driving simulators are better than their fixed-base counterparts (either these have hydraulic actuators or not). It seems that the most important advantage of moving-base simulators is the decrease of simulator sickness and better lateral control of the vehicle. The critical question here is if the research application areas of driving simulators can justify all this investment in moving-base systems. The author has to agree with Evans (1991) comment that at least in the aircraft industry, an aircraft simulator is "a 30 million dollar device representing a 150 million dollar aircraft. For the automobile case, it seems harder to justify a 30 million dollar simulator, when the real article can be purchased for about 10 thousand dollars". It is almost impossible to build a driving simulator which can be used for any research application area without exceeding the budget limits of the entrepreneur and more over to be valid for any type of application area. It should be always considered the fact that maybe a less sophisticated driving simulator could lead to the same valid results for a particular type of application.

Driving simulators have been built to ease, improve and promote traffic safety. As Evans (1991) stated "...traffic safety is one of many fields that can be characterised as data rich, understanding poor. The main thing that has been missing from traffic safety research is the appropriate scientific tradition to extract meaning from copious data that already exist; the answers to many key questions are embedded in existing data". Possibly, this is one of the reasons why the differences obtained from the simulated behavioural validation studies cannot be interpreted; the answers are already there, they are just obscured from the statistical details.

"In the era before the birth of experimental science, Greek philosophers thought that nature could be understood by pure thought alone, without the need for data. Nowadays there seem to be people who think that it can be understood with data alone, without the need for thought... Somewhat similar to the call for more data is the call for better driving simulators..." (Evans, 1991).

8. REFERENCES

AIKEN, L.R.(1971). *Psychological Testing and Assessment.* Badger, S. (eds). pp 87-115.

ALICANDRI, E., ROBERTS, K. and WALKER, J. (1986). A validation study of the DOT/FHWA highway simulator (HYSIM). *U.S. Department of Transport, Federal Highway Administration.* FHWA/RD-86/067.

ALLEN, R.W., KLEIN, R.H. and ZIEDMAN, K.(1979). Automobile research simulators: a review and new approaches. *Transportation Research Board* **706**, pp 9-15.

ALLEN, R.W., MITCHELL, D.G., STEIN, A.C. and HOGUE, J.R. (1991). Validation of real-time man-in-the-loop simulation. *VTI Raport.* No **372A**, Part 4, pp 18-31.

ALLEN, R.W. AND O' HANLON, J.F. (1979). Driver steering performance effects of roadway delineation and visibility conditions. *Paper presented at the 58th Annual Meeting of TRB*, Washington, D.C.

ALM, H. (1995). Driving simulators as research tools - a validation study based on the VTI driving simulator. Unpublished.

AMERICAN PSYCHOLOGICAL ASSOCIATION (APA), (1985). *Standards for Educational and Psychological Testing.* American Psychological Association (pubs). pp 9-23.

ANASTASI, A. (1988). *Psychological Testing.* Macmillan Publishing Company, New York.

BARRETT, G.V., NELSON, D.D., and KERBER, H.E. (1965). Human factors evaluation of a driving simulator. *Goodyear Aerospace Corporation* Final Report. Contract No PH 108-64-168.

BENSON, A.J. et al. (1989). Thresholds for the perception of whole body angular movement about a vehicle axis. *Aerospace Medical Association.* Washington, DC.

BERTOLLINI, G.P., JOHNSTON, C.M., KUIPER, J.W., KUKULA, J.C., KULCZYCKA, M.A. and THOMAS, W.E. (1994). The General Motors driving simulator. *SAE Technical Paper Series.* No **940179**.

BLAAUW, G.J. (1982). Driving experience and task demands in simulator and instrumented car: a validation study. *Human Factors* **24**(4), pp 473-486.

BLAIWES, A.S., PUIG, J.A. and REGAN, J.J. (1973). Transfer of training and the measurement of training effectiveness. *Human Factors* **15**(6), pp 523-533.

BLANA, E. (1994). *Comparison of Greek and English drivers' attitudes and behaviour regarding speed.* MSc Thesis. Institute for Transport Studies. University of Leeds. Leeds. UK.

BLANA, E. (1996). A survey of driving research simulators around the world. *ITS Working Paper* **481**. Institute for Transport Studies. University of Leeds. Leeds. UK.

BOULANGER, O. and CHEVENNEMENT, J. (1995). Analytical and application experiments: two necessary approaches for the driving simulator validity. *Paper presented at the Driving Simulator Conference.* TEKNEA. Toulouse, France. pp 26-39.

BREDA, W.M., KIRKPATRICK, M., and SHAFFER, C.L. (1972). A study of route guidance techniques. *North American Rockwell.* Final Report. Contract No. DOT-FH-11-7708.

BROWN, J.L. (1975). Visual elements in flight simulator. Report of Working Group 34. Washington DC: *Committee on Vision, Assembly of Behavioural and Social Sciences. National Research Council.* National Academy of Sciences.

CARO, P.W. (1973). Aircraft simulators and pilot training. *Human Factors* **15**(6), pp 502-509.

CIRILLO, J.A. (1968). Interstate system accident research study II, interim report II. *Public Roads* **35**, pp 71-75.

CRAWFORD, A. (1961). Fatigue and driving. *Ergonomics* **4**, pp 143-54.

CRONBACH, L.J. and MEEHL, P.E. (1955). Construct validity in psychological terms. *Psychological Bulletin* **52**, pp 281-302.

DROSDOL, J. and PANIK, F. (1985). The Daimler-Benz driving simulator: a tool for vehicle development. *SAE. The Engineering Resource For Advancing Mobility*. No **850334**. International Congress and Exposition, Detroit, Michigan.

DUNCAN, B. (1995). Calibration trials of TRL driving simulator. *Transport Research Laboratory*. Unpublished Project report. PA/3079/95. S221A/RB.

EBEL, R.L. (1961). Must all test be valid. *American Psychologist* **16**. October.

ELLIS, N.C. (1967). Using the Null Hypothesis in Human Engineering Evaluations. *Human Factors* **9**(4), pp 321-324.

EVANS, L. (1991). *Traffic Safety and the Driver*. Van Nostrand Reinhold (Publ.), New York.

FLEXMAN, R.E. and STARK, E.A. (1987). *Handbook of Human Factors.* Training simulators. *In*: G. Salvendy (Ed.). New York: Wiley & Sons. (pp 1012-38).

GHISELLI, E.E. (1959). The generalization of validity. *Personnel Psychology* **12**, pp 397-402.

GHISELLI, E.E. (1966). *The validity of occupational aptitude test.* New York: Wiley

GLEINMAN, (1991). *Psychology*. Third Edition. Norton International Student Edition. New York

GODTHELP, H., MILGRAM, P. and BLAAUW, G.J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors* **26**(3), pp 257-268.

HARMS, L. (1993). Driving Performance on a real road and in a driving simulator: results of a validation study. *Paper presented at the 5th International Conference on Vision in Vehicles*, Glasgow.

HARMS, L., ALM., H. and TÖRNOS, J. (1996). The influence of situation cues on simulated driving behaviour: a summary of three validation studies. *Paper presented at the Symposium on the Design and Validation of Driving Simulators.* ICCTP'96. Valencia.

HART, S.G. and STAVELAND, L.E. (1988). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research,* ref. *In* P.A. Hancock and N. Meshkati (eds.), Human Mental Workload, Elsvier Science Publishers B.V. (North-Holland).

HAUER, E. (1971). Accidents, overtaking and speed control. *Accident analysis & Prevention* **3**, pp 1-12.

HOFFMAN, E.R. and JOUBERT, P.N. (1966). The effect of changes in some vehicle-handling variables on driver steering performance. *Human Factors* **8**(3), pp 145-263.

HOGEMA, J.H., VAN DER HORST, A.R.A. and BAKKER, P.J. (1993). Chevron markings on the A59 and driving behaviour in fog (in Dutch). *Rapport IZF* **1993 C-9**. TNO Institute for Perception. Soesterberg. The Netherlands. (cited in Kaptein et al, 1995).

VAN DER HORST, A.R.A. (1990). *A time-based analysis of road user behaviour in normal and critical encounters*. Thesis. TNO Institute for Perception.

VAN DER HORST, A.R.A and HOEKSTRA, W. (1993). The perception of chevrons in fog: a simulator study (in Dutch). *Report IZF* **1993 C-10**. TNO Institute for Perception. Soesterberg. The Netherlands. (cited in Kaptein et al, 1995)

ISO TECHNICAL REPORT 3888 (1975). Road vehicles - Test procedure for a severe lane-change manoeuvre. *ISO*/**TR3888-1975(E)**.

JANSSEN, W.H. (1979). Routeplanning en geleiding: Een literatuurstudie. *Report IZF* **1979 C-13**. TNO Institute for Perception. Soesterberg. The Netherlands.

JANSSEN, W.H., VAN DER HORST, A.R.A. and HOEKSTRA, W. (1991). A simulator study into the effects of variable message signing on route choice and driving behaviour (in Dutch). *Report IZF* **1991 C-5**. TNO Institute for Perception. Soesterberg. The Netherlands. (cited in Kaptein et al, 1995).

JANSSEN, W.H., VAN DER HORST, A.R.A. and HOEKSTRA, W. (1992a). Descriptive information presentation on variable message signs (in Dutch). *Report IZF* **1992 C-7**. TNO Institute for Perception. Soesterberg. The Netherlands. (cited in Kaptein et al, 1995).

JANSSEN, W.H., VAN DER HORST, A.R.A. and HOEKSTRA, W. (1992b). Effects of variable message signing on route choice and driving behaviour: the influence of time loss and the presence of other traffic (in Dutch). *TNO Technische Menskunde*. Soesterberg. The Netherlands. (cited in Kaptein et al, 1995).

JOVANIS, P.P. (1995). Challenges to the design of driving simulator experiments. *Paper presented at the* Driving Simulator Conference. TEKNEA. Toulouse, France. pp 96-102.

KAPTEIN, N.A., THEEUWES, J. and VAN DER HORST R. (1995). Driving simulator validity: some considerations. *Pre-print of a paper presented at the TRB 75th Annual Meeting*.

KAPTEIN, N.A., VAN DER HORST R. and HOEKSTRA, W. (1996). The effect of field of view and scene content on the validity of a driving simulator for behavioural research. *Paper presented at the Symposium on the Design and Validation of Driving Simulators*. ICCTP'96. Valencia.

KAPPÉ, B. and KÖRTELING, J.E. (1995). Straight road steering: visual information in driving simulators (in Dutch). In preparation. *TNO Technische Menskunde*. Soesterberg. The Netherlands. (cited in Kaptein et al, 1995).

LEONARD, J.J. Jr and WIERWILLE, W.W. (1975). Human Performance Validation of Simulator; Theory and Experimental Verification. *Paper presented at the 19th Annual Meeting of Human Factor Society,* pp 446-455.

MACDONALD, W.A. and HOFFMANN, E.R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Human Factors* **22**(6), pp 733-739.

MALATERRE, G. (1995). Comparison between simulation and actual driving situations: some experiments. *Paper presented at the Driving Simulator Conference*. TEKNEA. Toulouse, France. pp 60-76.

MCCORMICK, E.J. (1970). *Human Factors Engineering*. McGraw-Hill. New York.

MCCOY, W.K. Jr. (1963). Problems of validity of measures used in investigating man-machine system. *Human Factors* **5**, pp 373-377.

MCLANE, R.C. and WIERWILLE, W.W. (1975). The influence of motion and audio cues on driver performance in an automobile simulator. *Human Factors* **17**, pp 488-501.

MCRUER, D.T., ALLEN R.W., WEIR, D.H. and KLEIN, R.H. (1976). New results in driver steering control models. *Human Factors* **19**, pp 381-397.

MCRUER, D.T., and KLEIN, R.H. (1975). Comparison of Driver Dynamics with Actual and Simulated Visual Displays. Paper No. **173**. *Systems Technology, Inc*. Hawthorne, California.

MCRUER, D.T., and KRENDEL, E.S. (1974). Mathematical models of human pilot behaviour. Paris: *NATO AGARD-AG-188.*

VAN DER MEDE, P.H.J. and VAN BERKUM, E.C. (1993). *The impact of traffic information: dynamics in route and departure time choice.* Thesis. Delft University of Technology, Delft, The Netherlands.

MESSICK (1980). Test validity and the ethics of assessment. *American Psychologist* **35**, pp 1012-1027.

MICHON, J.A. (1985). A critical review of driver behaviour models: what do we know, what should we do? *In*: L. Evans and R.C. Schwing (eds.). *Human behaviour and traffic safety*. New York: Plenum Press.

MORAAL, J. and POLL, K.J. (1979). De Link-Miles rijsimulator voor pantservoertuigen; verslag van een validitieonderzoek (The Link-Miles driver simulator for tracked vehicles; a validation study). *Report IZF* **1979-23**. TNO Institute for Perception. Soesterberg. The Netherlands.

MUDD, S. (1968). Assessment of the fidelity of dynamic driving simulators. *Human Factors* **10**(4), pp 351-358.

NÄÄTÄNEN, R. and SUMMALA, H. (1976). *Road-user Behaviour and Traffic Accidents.* Amsterdam, the Netherlands: North Holland.

NILSSON, L. (1989). The VTI driving simulator. Description of a research tool. *VTI Report* **150**.

PONDS, R.W.H., BROUWER, W.H., and VAN WOLFFELAAR, P.C. (1988). Age difference in divided attention in a simulated driving task. *Journal of Gerontology* **43**(6), pp 151-156.

PROVENMIRE, H.K., ROSCOE, S.N. (1973). Incremental transfer effectiveness of a ground-based general aviation trainer. *Human Factors* **15**(6), pp 534-542.

REED, M.P. and GREEN, P. (1995). Validation of a low-cost driving simulator using a telephone dialling task. *UMTRI*-95-19.

RIEMERSMA, J.B.J., VAN DER HORST, A.R.A. and HOEKSTRA, W. (1990). The validity of a driving simulator in evaluating speed-reducing measures. *Traffic Engineering + Control.* pp 416-420.

ROBERTS, K.M. (1980). The FHWA highway driving simulator. *Public Roads* **44**(3).

ROLFE, J.M., HAMMERTON-FRASE, A.M., POULTER, R.F., and SMITH, E.M.B. (1970). Pilot response in flight and simulated flight. *Ergonomics* **13**, pp 761-68.

SOLOMON, D. (1964). Accidents on main rural highways related to speed, driver and vehicle. Washington, *DC. Federal Highway Administration, US Department of Transportation*, July.

SOMA, H., HIRAMATSU, K., SATOH, K. and UNO, H. (1996). System architecture of the JARI driving simulator and its validation. *Paper presented at the Symposium on the Design and Validation of Driving Simulators.* ICCTP'96. Valencia.

STAPLIN, L. (1995). Simulator and field measures of driver age differences in left turn gap judgements. *Paper presented at the 74th TRB Annual Meeting.* Washington, DC.

TENKINK, E. (1988). Determinanten van rijsneheil (in Dutch). *Report IZF* **1988 C-3**. TNO Institute for Perception. Soesterberg. The Netherlands.

TENKINK, E. (1989). The effect of road width and obstacles on speed and course behaviour (in Dutch). *Report IZF* **1989 C-4**. TNO Institute for Perception. Soesterberg. The Netherlands.

TENKINK, E. (1990). Effects of road width and a secondary task on speed and lateral control in car driving (in Dutch). *Report IZF* **1990 C-27**. TNO Institute for Perception. Soesterberg. The Netherlands.

TENKINK, E. and VAN DER HORST, A.R.A. (1991). Effects of road width and curve characteristics on driving speed (in Dutch). *Report IZF* **1991 C-26**. TNO Institute for Perception. Soesterberg. The Netherlands.

TIFFIN, J. and MCCORMICK, E.J. (1965). *Industrial Psychology*. Englewood Cliffis. New Jersey: Prentice Hall.

VALVERDE, H.H. (1973). A review of flight simulator transfer of training studies. *Human Factors* **15**(6), pp 510-523.

WHEATON, G.R., KINSLOW, W.E., and KRUMM, R.L. (1966). Validation of a part-task automobile driving simulator. *Radio Corporation of America* Contract No. PH 108-64-167.

WILLIGES, B.H., ROSCOE, S.N. and WILLIGES, R.C. (1973). Synthetic flight training revisited. *Human Factors* **15**(6), pp 543-560.

WINER, B.J. (1962). *Statistical Principles in Experimental Design*. New York. McGraw-Hill Book Company, Inc.

WOJCIK, C.K. and WEIR, D.H. (1970). Studies of the driver as a control element -phase 2. Report No 70-j73. *Institute of Transportation and Traffic Engineering*. Uni. of California. L.A.

9. APPENDIX A

List of abbreviations for Table 9-1

Column 1: Names of the researchers and year conducted the validation studies

Column 2: Name and type of driving simulators

Name:
UMTRI: University of Michigan Transport Research Institute driving simulator, USA
TRL: Transport Research Laboratory driving simulator, England
TNO: Institute for Perception, Soesterberg, the Netherlands
INRETS: France
DAIMLER-BENZ: Germany
VTI: Linköping, Sweden
FHWA: Federal Highway Administration Highway Driving Simulator (HYSIM), USA
JARI: Japan

Type:
FB: Fixed-Base
FB_H: Fixed-Base with Hydraulic actuators (i.e. limited motion system)
MB: Moving-Base

Column 3: Data collection technique and training sessions

There are three main ways of collecting real road data:
1) using an Instrumented Vehicle (IN.VE.) on the real road (RR)
2) using an Instrumented Vehicle (IN.VE.) on a Test Track (TT)
3) using Genuine Road Users (GRU)

Training sessions indicate the training of subjects before the real and/or the simulated experiment

Column 4: No of subjects used on the real road and simulated experiments.

This column also includes information about the gender, age, no of years holding the driving licence, no of kilometres driven per year (driving experience).
M: Male
DL: No of years holding the Driving Licence
DD: Driving Distance in km or miles (the American and English studies are using the imperial system)
y: year
A: Age
AA: Average Age
ADY: Average Driving Years
INE: INExperienced drivers
EX: EXperienced drivers

Column 5: Type of statistical analysis used to compare the real road and the simulated data

CoM: Comparison of Means using paired-t and/or unpaired-t tests
ANOVA: ANalysis Of VAriance
NASA-TLX: test for measuring the metal workload
PROC GLM: General Linear Models Procedure in SAS with program options for repeated-measures designs, within each block of trials corresponding to a single experimental methodology.
PROC CORR NOMISS: Correlation Procedure in SAS with program option for not missing cases
normalised amplitude density functions
Pearson product-moment correlation= linear correlation between two variables
Newman-Keuls tests
time series analysis

Column 6: Independent Variables

The independent variables have not been abbreviated

Column 7: Dependent Variables

Study Blaauw, 1982
M_LP INE.= Mean Lateral Position of INExperienced drivers
M_LP EX.= Mean Lateral Position of EXperienced drivers
Var_LP INE.= Variance of Lateral Position of INExperienced drivers
Var_ LP EX.= Variance of Lateral Position of EXperienced drivers
SD_SWA INE.= Standard Deviation of Steering-Wheel Angle of INExperienced drivers
SD_SWA EX.= Standard Deviation of Steering-Wheel Angle of EXperienced drivers
SD_yaw rate INE.= Standard Deviation of yaw rate of INExperienced drivers
SD_yaw rate EX.= Standard Deviation of yaw rate of EXperienced drivers
M_S INE.= Mean Speed of INExperienced drivers
M_S EX.= Mean Speed of EXperienced drivers
SD_S INE.= Standard Deviation of Speed of INExperienced drivers
SD_S EX.= Standard Deviation of Speed of EXperienced drivers
SD_ACC INE.= Standard Deviation of Acceleration of INExperienced drivers
SD_ACC EX.= Standard Deviation of Acceleration of EXperienced drivers

Study Alicandri
A. SI. DE.D.= Average SIgn DEtection Distance
SD_DE.D.= Standard Deviation of DEtection Distance (the point at which a subject first reported seeing a sign of a specific background colour stated by the experimenter)
A. SI. REC. D.= Average SIgn RECognition Distance (the distance form a sign at which the subject could read/understand the sign)
SD_REC. D.= Standard Deviation of RECognition Distance
S= Speed
A.P.C.= Accelerator Position Changes (any change in the position of the pedal incorporating ≥10 percent of its total travel)
Steering Wheel Reversals (any movement of steering wheel exceeding 20 degrees)

Study Riemersma et al, 1990
Vavg= average entrance speed in km/h only for the last run (the twelfth run)
SD_Vavg= Standard Deviation of the average speed
the confidence intervals are in km/h

Study Hogema, 1992
M_ratings = Mean of ratings; high ratings indicate an easy driving task
SD_ratings = Standard Deviation of ratings
M_cone displ.= Mean of cone displacement
SD_cone displ. = Standard Deviation of cone displacement
M_SRR= Mean of Steering Reversal Rate
SD_SRR = Standard Deviation of Steering Reversal Rate
max. safe speed= the highest speed at which the driver could complete the lane change manoeuvre both safely and successfully. The criterion for safety was that subjects had to keep the vehicle in control the whole time whereas for success was that in two subsequent changes at a certain speed no more than 2 cones were displaced in each run.
SRR= the number of times per minute that the direction of the steering wheel movement is reversed through a small angle or gap (MacDonald and Hoffmann, 1980).

Study Harms, 1993; Alm, 1995; Harms et al, 1996 (VTI validations)
M_S=Mean Speed in km/h
SD_S= Standard Deviation of Speed
M_LP= Mean Lateral Position from the real left wheel of the vehicle to the centreline of the road in cm
SD_LP= Standard Deviation of Lateral Position
LP (Harms et al)= Mean Lateral Position from the centre of the vehicle to the centreline of the road
The "on" and "off" means the moving system was either on or off.
The "with" and "without" means that the subjects could either have access to the speedometer or not.

Study Malaterre, 1995
TH= Time Headway (following distance/speed)
HR= Heart Rate (no of pulses per minute)
SRR= Steering Reversal Rate (no of steering reversal movements per minute)

Study Green and Reed
M_LP= Mean Lane Position positive to the right of the centreline (ft from left edge). The data were collected using two lane-tracking cameras mounted in the side mirrors and aimed at the road.
SD_LP= Standard Deviation of Lateral Position (ft)
M_LS= Mean Lateral Speed, i.e. mean of absolute values of first-order differences in lane position (ft/s)
M_LTL= Mean Log Time to Line (log seconds). Time-to-line was calculated for each data interval (30/sec) as the time that would be required for a wheel of the car to reach a lane-edge marker if the lateral velocity remained constant at the value calculated from the first-order difference in lane position (Godthelp et al, 1984).
SD_S= Standard Deviation of Speed (mile/h)
M_AA= Mean Absolute Acceleration, i.e. mean of absolute values of first-order differences in speed (mile/h per second)
SD_ SWA= Standard Deviation of Steering Wheel Angle (degrees)

SRF= Steering Reversal Frequency (1/s), i.e. mean number of steering-wheel reversals per second. a discrete SW motion consists of a series of first-order SWA difference that do not change sign for more than 0.33 second (ten samples at 30 Hz) and that represent a net monotonic change in SW position of more than 1 degree. The number of reversals during a trial is one less that the number of discrete motions. Dividing by the duration of the trial give the SWRs.
SD_TP= Standard Deviation of Throttle Position (% of full throttle)
TRF= Throttle Reversal Frequency (1/s). It is calculated similarly to the SWF, except that a discrete throttle motion consists of a net monotonic change of 0.5 percent with a duration of at least 0.33 seconds.

Study Duncan
M_S=Mean Speed in mile/h
SD_LP= Standard Deviation of Lateral Position
SD_S (anti)= Standard Deviation of mean Speed anticlockwise
n.s.d.= no significant difference
s.d.= significant difference
MC SH= Mean Choice of Safe Headway
M_FD= Mean Following Distance for the "fixed" headway circuits (target headway set at 30m)
M_ glances= Mean number of glance
M_ min H= Mean of the minimum Headway

Study Staplin
TRD= Target Recognition Distance (only for the real road experiment)
"LSG"D= "Least Safe Gap" Distance

Study Soma et al, 1996
Va= average Velocity of vehicle in
TP= Time required Passing the course
Vp-p= difference between maximum and minimum velocity of vehicle in
APmax= maximum open ratio of Acceleration Pedal
TRA= Time Ration of putting on the Acceleration pedal
SWAd= differential of Steering Wheel Angle
Yrd= differential of Yaw velocity
SWA1-6= first to sixth peak of Steering Wheel Angle
YR1-6= first to sixth peak of Yaw velocity
LA1-6= first to sixth peak of Lateral Acceleration

Study Kaptein et al, 1996
M_TTCbr= mean Time-To-Collision at the onset of the braking manoeuvre in sec
TTCmin= the minimal TTC that occurred during each braking manoeuvre in sec
ACCmin= minimal ACCeleration in
DISTstop= distance from the other vehicle when stopped in
DISTbr= distance to the stationery vehicle at the onset of braking in

Column 8: Results

The results are given in mean values which correspond in Column 7 dependent variables. Please read horizontally columns 7 and 8 were applicable.

Column 9: Comments

Comments relevant to the particular validation study are given in this column.

Validation studies	Driving simulator (type of d.s.)	Data collection technique	No of subjects	Statistical analysis	Independent variables	Dependent Variables	Results (mean values)			Comments
							SIM	CAR	COR	
Blaauw (1982)	TNO (old version), FB	IN.VE. RR, **no** training sessions	48 M (24EX +24 INE) EX: DL:>3y DD:30000 km INE: DL: train. course or just passed the DL	mean values, SD, normalised amplitude density functions, Pearson product-moment correlations, ANOVA, Newman-Keuls tests	driving instructions (free driving, forced lateral control, forced long. control, forced lat. and long. control), two driving conditions, sequence of the two conditions	LP INE. LP EX Var_ LP INE. Var_LP EX. SD_SWA INE. SD_SWA EX. SD_yaw rate INE SD_ yaw rate EX. S INE. S EX. SD_S INE. SD_S EX. SD_ACC INE. SD_ACC EX.	190.6 194.2 36.4 24.3 1.8 1.2 0.31 0.26 104.9 103.4 1.3 1.0 2.3 1.3	178.4 171.2 19.4 16.6 1.6 1.4 0.32 0.32 104.3 109.7 1.1 0.8 1.4 1.1	0.36 (all) 0.57 (all) 0.14 (all) 0.32 (all)	straight road section; good abs. and relat. validity for long. veh. control and good relative validity for lateral veh. control
Alicandri, Roberts, Walker (1986)	HYSIM, FB, night time	IN.VE., RR training sessions (10 min field+ sim)	32 (16M+16F) AA=32.4, ADY=14.8	5-way ANOVA; PROC GML (DE., REC., & APC); PROC ANOVA (S, SWR); PROC CORR NOMISS	(sex, age), the two driving conditions, type of signs (4), the data collection zone	A. SI. DE.D SD_DE.D A. SI. REC. D. SD_REC. D. S A.P.C. SWR	498	659		500ft visual range for small road 900ft for highways; physiological meas.

Validation studies	Driving simulator (type of d.s.)	Data collection technique	No of subjects	Statistical analysis	Independent variables	Dependent Variables	Results (mean values)				Comments
							SIM		CAR	COR	
Riemersma et al (1990)	DAIMLER-BENZ, MB	RR, GRU, training session	24 M A: 20-35y (Sim), previously driven the simulator	ANOVA	the two driving conditions, driving instructions (calm, fast)	**before** **Run 1/2** Vavg SD_Vavg 95% conf. interval **Run 8/9** Vavg SD_Vavg 95% conf. interval **Run 12** Vavg SD_Vavg 95% conf. interval **after** **Run 1/2** Vavg SD_Vavg 95% conf. interval **Run 8/9** Vavg SD_Vavg 95% conf. interval **Run 12** Vavg SD_Vavg 95% conf. interval **change (%)** **Run 1/2** **Run 8/9** **Run 12**	calm 75.8 8.1 70.4-81.2 64.0 8.1 58.6-69.4 61.6 5.4 55.4-67.8 53.3 16.8 44.9-61.7 54.8 6.1 50.7-58.9 53.3 8.2 43.8-62.6 29.7 14.3 13.6	fast 103.3 16.8 92.1-114.4 87.9 17.8 76.1-99.7 77.2 12.0 63.3-91.0 62.6 28.4 43.8-81.5 65.1 13.2 56.3-73.9 60.0 7.5 51.3-68.7 39.4 25.9 22.3	 72.4 11.9 70.0-74.8 72.4 11.9 70.0-74.8 72.4 11.9 70.0-74.8 66.2 11.9 63.8-68.6 66.2 11.9 63.8-68.6 66.2 11.9 63.8-68.6 8.6 8.6 8.6		traffic calming, rural village

Validation studies	Driving simulator (type of d.s.)	Data collection technique	No of subjects	Statistical analysis	Independent variables	Dependent Variables	Results (mean values)				Comments
							SIM		CAR	COR	
Hogema (1992)	TNO, FB	IN.VE., TT	12 M DD:>30000 km, previously driven the simulator	CoM, ANOVA	comp. and uncompensated driv. sim. and IN.VE; 3 fixed speeds (30, 45, 60 km/h)	M_ratings SD_ratings M_cone displ. SD_cone displ. M_SRR SD_SRR max. safe speed	uncom 4.96 2.44 1.74 1.93 51.4 19.8 55	comp 5.24 2.46 1.37 1.65 47.0 15.3 58	 7.36 2.14 0.04 0.23 36.6 9.4 72-75		compens. technique relative valid for SRR, no absolute valid for the double lane task physiological meas.
Harms (1993)	VTI MB	IN.VE., RR, 2 train sessions,	7	ANOVA, correl., CoM	the two driving conditions, driving sessions 3+3 sessions S&RR, 8 trials	M_S SD_S M_LP (cm) SD_LP	81.7 70.6		79.0 92.4	0.87 0.49	no oppos. traffic, no road envir.
Alm (1995)	VTI MB	IN.VE., RR, 20 Km train. session	17	one-way ANOVA, correl., CoM, NASA-TLX	the two driving conditions, moving base on-off	M_S on SD_S on M_S off SD_S off M_LP on (cm) SD_LP on M_LP off (cm) SD_LP off	84.2 8.51 85.1 8.31 78 32 76 42		83.9 6.23 83.9 6.23 73 29 73 20		

Validation studies	Driving simulator (type of d.s.)	Data collection technique	No of subjects	Statistical analysis	Independent variables	Dependent Variables	Results (mean values)			Comments
							SIM	CAR	COR	
Harms et al (1996)	VTI MB	IN.VE, tunnel	20	one-way ANOVA, correl., CoM,	access to speedometer; position of the tunnel wall (TW); the two driving conditions	M_S with M_S without LP TW left LP TW right	81.0 84.7 1.9 1.5	73.4 75.4 2.1 1.7		
Malaterre (1995)	INRETS FB	IN.VE., TT	8	CoM, questionnaire	Driving instructions (comfort and min distance), and 4 dif. speeds; the two driving conditions	TH, HR, SRR				
Reed and Green (1995)	UMTRI FB	IN.VE., RR training sessions	12 (6M+6F) A:20-30y & >60y	ANOVA	age (young, old), gender, fidelity (high, low, normal), task (phone, normal)	M_LP SD_LP M_LS M_LTL SD_S M_AA SD_SWA SRF SD_TP TRF	5.32 1.18 0.45 1.26 0.76 0.14 2.44 0.31 1.44 0.07	6.39 0.54 0.16 1.65 0.89 0.22 0.58 0.08 2.77 0.23	0.43 0.59 0.76 0.58 0.19 0.34 0.74 0.48 0.45 0.18	low cost driving simulator

Validation studies	Driving simulator (type of d.s.)	Data collection technique	No of subjects	Statistical analysis	Independent variables	Dependent Variables	Results (mean values)			Comments
							SIM	CAR	COR	
Duncan (1995)	TRL FB_H	IN.VE., TT 15min training	47 (30M & 17F, AA:40.5M & 42.5F), DD:2000m py, DL:>2y,	Paired t-tests, average data for primary tasks, unpaired t-tasks for secondary tasks	1)order effects both driving conditions 2)clockwise/anticlockwise both driving conditions 3)lane keeping +in-car distrac 4) eye glance 5) headway	M_S SD_LP SD_LP SD_ S(anti) SD_LP lateral deviation lateral deviation M_ glances eye-off road time glances on display MC SH M_ FD M_ min H	+0.72 mile/h n.s.d. 2*X -1.12 mph (s.d) +22% -10% (s.d) +22% (s.d.) 20.1 16.4 secs 16.5 62.8m 64.2m 41.6m	+1.3 n.s.d. X -0.23 +28% -18% (s.d) +6% (n.s.d.) 23.8 19.1 17.9 50.8 45.1m 30.0 m		drivers weaved more in the simulator
Staplin (1995)		IN.VE., TT	79(47M+32 F) (i:25+ii: 29+iii:25) AA: (i:33.3, ii: 65.1, iii: 79.4)	PROC GLM	3 types of visual display, two target speeds (30 and 60 mile/h), three groups of ages (i, ii, iii); the two driving conditions	TRD, "LSG"D				

Validation studies	Driving simulator (type of d.s.)	Data collection technique	No of subjects	Statistical analysis	Independent variables	Dependent Variables	Results (mean values)			Comments
							SIM	CAR	COR	
Soma et al (1996)	JARI MB	IN.VE., TT Training sessions	11 M A:>65y	correlation time series analysis	Moving system on and off; the two driving conditions	Va, TP, Vpp, Apmax, TRA, SWAd, Yrd, SWA1-6, YR1-6, LA1-6 % good speed control (40±5 km/h)	76.9on and 71.8 off	77.8		
Kaptein et al (1996)	TNO, FB (updated version)	IN.VE., TT training sessions	12 M DL>5y DD>10000 km/y A: 25-45; 3 sessions	ANOVA t-test	Field of view (40o and 120o); scene complexity (plain and complex); target speeds (30, 60, 90 and 120 km/h); normal and hard braking; the two driving conditions	M_TTCbr TTCmim ACCmin DISTstop DISTbr				side winds were simulated

Table 9-1 Summary of driving simulators behavioural validation studies