Calibration in Young Drivers

Young drivers are often designated as a problem group. There are indications that they have relatively poor risk recognition and that they underestimate what is needed to handle potentially dangerous situations. At the same time they tend to overestimate their own driving skills. Calibration is the process of tuning task difficulties and one's skills; the ability to balance task demands and skills accurately.

In the research programme 'Calibration in young drivers' empirical research is conducted into the process of calibration during driving. The final goal of this research is to make recommendations for driver education or a supplementary (second phase) course or feedback to young drivers. This paper discusses a pilot-study with the objective to assess the methods used to measure calibration and to further develop these methods. A questionnaire was tested in a pre- post-test setting with 20 inexperienced drivers, who all attended training. The training was not specifically aimed at improving calibration; it was a regular post-licensing training.

The tests on reliability and different types of validity suggest that, after some adaptations, the questionnaire is a sufficient instrument to measure calibration. A high reliability, or internal consistency, of the constructs was found. And there was enough evidence for construct-validity, although, the questionnaire was adjusted somewhat as a result of the tests.

Only the construct-related validity was disappointing, because no association could be established between the self-reported calibration scores and the direct ratings on calibration given by experts. However, several causes are suggested to explain the absence of this correlation.

Introduction

Young drivers are often designated as a problem group. Accident data, for example, show that traffic accidents continue to be the leading cause of death for people aged 15 to 24 in all western motorised countries. Studies of behaviour and accidents show errors that are often made by young drivers in particular. One of these beginners' errors is the wrong assessment of the demands of the driving task. Danger and risk are not recognized as such (or are recognized too late). This can be a result of insufficient exploring the environment and looking around. But it can also be caused by a deviant acceptation criterion (for example the perceived safety-margin). In brief, there are indications that young drivers in particular have poor risk recognition.

At the same time there seems to be a problem with the assessment or evaluation of one's one driving skills. For example, the young driver underestimates what is needed to cope with a dangerous situation. This could be caused by the fact that he has encountered only a limited number of traffic situations, which may provide a false sense of mastery and safety. Kuiken and Twisk (2001) assume a direct relationship between self-assessment (the estimation of one's abilities) and the recognition and appraisal of danger. Because of a lacking awareness of one's skills and the dynamics of the traffic system, there is insufficient anticipation of risk and

dangerous situations. Calibration is the process of matching task difficulties and one's skills: the ability to balance task demands and skills accurately.

In the research programme 'Calibration in young drivers' empirical research is conducted into the process of calibration during driving. The final goal of this research programme is to make recommendations for driver education or a supplementary (second phase) course to young drivers. The first step in the research programme is the evaluation of a second phase training which improves self-assessment skills, risk and safety awareness. This behavioural training takes place on a closed circuit. The emphasis in the training lies on the perception of danger and on the cognitive aspects of driving (risk and safety awareness). One of the instruments to measure the effects of the second phase training is a questionnaire on self-assessment of risk and safety awareness. This questionnaire will be used in following phases of the research programme as well. It will be used to follow developments of young drivers during a period of four years. Finally, it will also be used in a study in a driving simulator, with more controlled conditions.

This questionnaire was tested in a pilot-study to discover if it is a valid and reliable instrument to measure calibration. The results of this pilot will be used to further develop the instrument.

Questionnaire used in the pilot study

1. Method

Subjects
The questionnaire was tested on 20 learner drivers who were at the end of their driver education. Half of the subjects were male. The average age of the participants was 24.

Instrument
The questionnaire which was used in the pilot is an adaptation of a questionnaire developed by Hatakka (1998). Beside some general questions, the questionnaire consists of four parts, to measure four different constructs:

1. Estimated strengths and weaknesses as a driver. Both skill-oriented (perceptual motor skills) and safety-oriented items.
2. Estimated danger and risks (task difficulty) concerning external factors. (i.e., difficult situations and behaviour of others)
3. Estimated danger and risks (task difficulty) concerning internal factors. (i.e., the drivers' own characteristics and habits)
4. Estimated strengths and weaknesses compared to other, more experienced, drivers.

Note that these constructs themselves do not tell us anything about calibration. Only a combination of answers on the estimated strengths and weaknesses (construct 1 and 4) and the answers on estimated danger and risks (construct 2 and 3) reveal something about the calibration skills.

In addition to this self-reported behaviour, an expert opinion on the calibration shown was asked from the driving instructor, who knew the students' driving ability inside out. The driving instructor first received an explanation of the concept of calibration and was then asked to directly rate the calibration-skills of their pupils on a scale of one to ten, where:
1 = "This pupil has a very low score on calibration (for example: doesn't recognize the dangers in traffic and strongly overestimates himself)"

5 = "This pupil has a mediocre score on calibration: sometimes calibration is quite well, but often it is insufficient"

10 = " This pupil scores very well on calibration: he has a very good idea of his own capabilities and shortcomings, he is aware of possible dangers in traffic and of the demands traffic and driving make, and he tunes his behaviour to the demands"

This expert opinion was asked so that it could be compared to the self-assessed calibration skills. Similar calibration-scores from different sources strengthen the idea that the instruments capture calibration.

Design

The design that was used in the pilot was a *within-subjects-without-control* design. The participants first completed the questionnaire on paper and then received a safety training. After the training the participants where asked to complete the written questionnaire a second time. A control condition was not necessary in this study, because the primary goal was not to scientifically assess the effects of the intervention. The intervention was a regular safety training, which was not aimed at improving calibration. However, it was a good opportunity to test the questionnaire in a setting which is comparable to the setting in the main experiment.

2. Results: Reliability and Validity

To assess the reliability and validity of the questionnaire, the results of the pre- and post-test were combined. The intervention does not affect the reliability or validity of the instrument. The only consequence of this adding up is that respondents are represented twice in the data, which is a violation of the independency assumption. Nevertheless, it was decided to use the combined database, as some of the respondents did not finish the pre-test and others did not finish the post-test. This means that in the combined database only some measurements are from the same respondents. The results of the reliability- and validity-tests are used only indicatively.

Reliability

The most important objective of the pilot-study was to legitimate the questionnaire. Therfore the first aspect we looked at was if the items formulated to measure each construct, were 'reliable' (Walsh and Betz, 1995).

Reliability involves the extent to which we measure a particular attribute in a systematic and therefore repeatable way. A reliable instrument can be trusted to provide the same score every time the same attribute is measured. A Cronbach's α was calculated for each construct (*Table 1*), with the combined scores on the pre- and post-test.

Table 1. Cronbach's α of the constructs in the questionnaire

	Cronbach's α
1: Estimated strengths and weaknesses as a driver. Both skill-oriented (perceptual motor skills) and safety-oriented items.	.82
2: Estimated danger and risks (task difficulty) concerning external factors. Difficult situations and behaviour of others.	.87
3: Estimated danger and risks (task difficulty) concerning internal factors. The drivers own characteristics and habits.	.94
4: Estimated strengths and weaknesses	

compared to other, more experienced, drivers.	.93

As can be seen in the table every construct has a Cronbach's α above .82. This indicates high reliability and internal consistency.

Criterion-related validity

The term criterion-related validity refers to the extent to which a measure or attribute shows an association with some independent or external indicator of the same attribute (Walsh & Betz, 1995). In this pilot study such an external indicator was available, namely the expert opinions on calibration capacities of the participants.

In order to calculate the association between the self-reported calibration and the opinion of experts, the respondents were divided into three groups, on the basis of their own opinion on danger and risks in traffic and their own assessment of qualities and weaknesses of their driving skill. The first group, the 'Dangerous mis-calibration' group, consisted of respondents who thought there is not much danger and risk (either from external or internal factors) in traffic and they were most convinced of their own driving abilities. The second group, 'Poor mis-calibration', consisted of those who indicated the most danger and risks in traffic, and were at the same time insecure about the driving skills. The rest of the respondents formed the third group ('Good calibration').

With these groups the following cross-tabulations were made (*Tables 2* and *3*). The cross-tabulations show the scores of the experts (there were only ratings above 4) and which calibration-group the pupil was assigned to according to their own calibration-score. The questionnaires and expert opinions on pre- and post-test were combined for this purpose. The displayed association statistic in *Tables 2* and *3* is the Eta measure of association.

Table 2. Comparison between self-assessment and expert opinion. The calibration groups were formed based on reported danger and risks concerning external factors.

r = .087		Rating by experts on calibration of the participants (1=very poor… 10=excellent)					
		5	6	7	8	9	Total
Danger and risks concerning *external* factors	Dangerous mis-calibration	1	1	1	3		6
	Poor mis-calibration		1	2			3
	Good calibration	3	1	2	4	1	11
	Total	4	3	5	7	1	20

Table 3. Comparison between self-assessment and expert opinion. The calibration groups were formed based on reported danger and risks concerning internal factors.

r = .246		Rating by experts on calibration of the participants (1=very poor… 10=excellent)					
		5	6	7	8	9	Total
Danger and risks concerning *internal* factors	Dangerous mis-calibration	1		1			2
	Poor mis-calibration		1	2	1		4
	Good calibration	3	2	2	6	1	14
	Total	4	3	5	7	1	20

There appears to be no association between the self-assessment of calibration based on the danger and risks of external factors and the expert opinion. Moreover, there is only a moderate association between calibration based on the danger and risks of internal factors and the expert opinion. From these cross-tabulations the conclusion was drawn that there is no evidence for criterion-related validity. There can be several causes for this absence of correspondence between the expert-opinion and the self-assessment. The explanation of calibration and the instruction for the driving instructor may have been unclear. Only half of the driving instructors answered the question, indicating there could be something wrong with the question. On the other hand the classification of pupils to several 'calibration-groups' is just an explorative one. The right way to obtain a calibration-score from the questionnaire will be explored further in the research programme.

Construct Validity

Construct validity addresses the question: "Does the test/questionnaire measure what it is intended to measure?" (Walsh & Betz, 1995). There are two indicators for construct validity. The first, convergent validity, applies when items measuring the same construct are (highly) correlated. The second indicator is discriminant validity, which is applicable when items measuring different constructs or traits are not correlated.

In order to assess the discriminant and convergent validity, a correlation table was made for each item of every construct. The answers to the pre- and post-test questionnaire were again combined for this purpose. The correlation table is schematically reflected in *Figure 1*. In this figure every row and every column represents an item, every box represents a correlation coefficient. The correlation coefficients that were found to be significant (α=5%) are displayed as a shaded box. All the significant correlation coefficients were positive; respondents who score high on one item also score high on the other item.

The existence of convergent validity can be found between the double-lines in *Figure 1*. The items in each construct seem to be correlated. This was to be expected, since the Cronbach's α was also very high for these constructs. When looking at the correlations of the first construct, "Estimated strengths and weaknesses as a driver", there seems to be a difference between the first eight items, and the last five items. This is illustrated in *Figure 1* with dashed lines. In the questionnaire, the first eight items all turn out to be related to driving skill ("controlling the vehicle", "react fast"), the last five items concern safety issues ("driving according to traffic laws", "driving very carefully"). Also in the fourth construct, "Estimated strengths and weaknesses compared to other, more experienced, drivers", there seems to be a difference between the first three items and the last ten items. This is illustrated in the figure with dashed lines, and corresponded with the distinction in the questionnaire between safety issues (the first three items), and driving skill (the last ten items). These distinctions between safety related and driving skill related items indicate that there may be a difference between estimated strengths and weaknesses in driving skill and safety. That is, someone can think of himself or herself as driving very skilfully (controlling the vehicle, reacting fast), but then state that he or she does not drive very safely. Driving skilfully and driving safely are not the same in the eye of the respondents. It looks like these are actually two different constructs.

Discriminant validity states that no correlation should exist between items that measure different constructs (traits). In the correlation tables this is reflected in the space outside the double lines. The pattern of correlations suggests that constructs 2 and 3 (danger and risks concerning internal and external factors) are actually one construct. Drivers do not make a distinction between internal and external factors in the estimation of danger and risks in traffic. The same holds for constructs 1 and 4, which also have many correlated items. The estimated strengths and weaknesses as a driver are the same as when they compare themselves with

other, more experienced, drivers. But this is only true for the items concerning driving skills. The association is less strong with the safety-oriented items. The respondents estimate their strengths and weaknesses in driving safely differently when they compare themselves with other, more experienced, drivers.

Construct 1) Estimated strengths and weaknesses as a driver
Construct 2) Danger and risks concerning external factors
Construct 3) Danger and risks concerning internal factors
Construct 4) Strengths and weaknesses compared to other, more experienced, drivers
Construct 1) Estimated strengths and weaknesses as a driver
Construct 2) Danger and risks concerning external factors
Construct 3) Danger and risks concerning internal factors
Construct 4) Strengths and weaknesses compared to other, more experienced, drivers

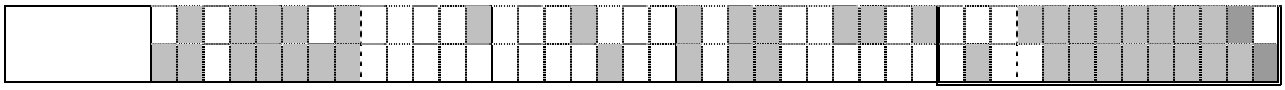

Figure 1. Schematic representation of the correlation table of all questions in the questionnaire.

3. Discussion and conclusions

The tests on reliability and different types of validity suggest that, after some adaptations, the questionnaire is a satifactory instrument to measure calibration with. Every construct has a Cronbach's α above .82, which indicates high reliability or internal consistency.

With regard to criterion-related validity no association was found between the self-assessment of calibration based on the danger and risks of external factors and the expert opinion. And only a moderate association was found between calibration based on the danger and risks of internal factors and the expert opinion. From these cross-tabulations the conclusion was drawn that there was no evidence for criterion-related validity. However, several disrupting circumstances may have occurred. For instance, the explanation of calibration and the instructions for the driving instructor may have been unclear. Only half of the driving instructors answered the question, indicating there may have been something wrong with the question. Also, the classification of pupils to several 'calibration-groups' was just an explorative one. The right way to obtain a calibration-score from the questionnaire will be explored further in the research programme.

There was however evidence for construct validity. The items within one construct did have high correlations, which indicates convergent validity. And the other requirement for construct validity, discriminant validity, was also met. The items measuring different constructs generally did not correlate. With two exceptions though; the two constructs measuring danger and risks concerning internal and external factors, were found to be actually one and the same construct. The respondents did not respond differently to internal and external factors of danger and risks. The self-reported strengths and weaknesses turned out to be also one construct, whether or not the comparison was made with other, more experienced, drivers. This association was less strong with the safety-oriented items, however. The respondents estimated their strengths and weaknesses concerning driving safely differently when they compared themselves with other, more experienced, drivers.

The questionnaire was adjusted to these findings. Constructs 2 and 3, internal and external factors of danger and risks, where combined to form one construct. And constructs 1 and 4 were also combined, but the contrast with and without the comparison to other drivers was maintained.

References

Hatakka, M. (1998). Novice Drivers' Risk- and Self-evaluations; Use of Questionnaires in Traffic Psychological Research. Method Development, General Trends in Four Sample Material, and Connection with Behaviour. Unpublished doctoral dissertation, University of Turku, Turku, Finland. Bieb: St 20000072.

Kuiken, M., & Twisk, D. (2001). (Report No. R-2001-29). Leidschendam, The Netherlands: SWOV Institute for Road Safety Research.

Walsh, W. B., & Betz, N. E. (1995). Tests and Assessment (Third ed.). Englewood Cliffs, New Jersey: Prentice Hall, Inc.