SELF-ASSESSMENT OF KNOWLEDGE AND ABILITIES

A literature study

Anna Sundström

EM No 54, 2005

ISSN 1103-2685
Self-assessment of knowledge and abilities
– a literature study

Anna Sundström

Driver education in some countries has recently come to include drivers’ risk awareness and evaluation of their own knowledge and abilities instead of focusing only on drivers’ knowledge and abilities, measured by tests or expert judgements. Drivers’ self-assessment has also been in focus in a new curriculum for driver education in Sweden. When this new curriculum is introduced, the driving-license test will have to be altered according to the new goals of the curriculum and new strategies for measuring drivers’ self-assessment have to be employed. There is limited knowledge of how self-assessment of driver competence should be measured. Therefore, the overall purpose of this literature study was to examine how self-assessment in general is measured in different fields. In addition, the purpose was to investigate how self-assessments of driver competence have been measured and also whether the methods used for self-assessment in general and for self-assessment of driver competence could be applied to the driver education system in order to measure the goals of the curriculum.

Studies about self-assessment of abilities and knowledge were reviewed. Since self-assessments might be slightly unreliable because of the respondents’ willingness to answer in a socially desirable way, the self-assessments could be compared to external measures. The conclusion was that the studies reviewed showed a homogenous picture. In general, the studies showed that there is a relationship between self-assessments and external measures, but the strength of the relationship varies between different studies. In addition, the relationship between self-assessment and external measures is affected by various factors. One such factor is the competence of the test-taker. Competent test-takers are more likely to make a realistic self-assessment than less competent test-takers. Other factors that affect the relationship are the specificity of the domains assessed, the difficulty of the domain assessed and if a relative or absolute judgement is made. Studies have also shown that the accuracy of self-assessment could be improved by different interventions such as education and observation of one’s own performance.

When examining studies of self-assessed driver competence and comparing them to studies of self-assessment in general, the conclusion was that self-assessments of driver competence have rarely been related to an external measure. In order to check the validity of drivers’ self-assessments, further studies are needed in order to relate self-assessments to an external measure. There is also a need for further studies in order to investigate if the factors affecting the accuracy of self-assessment also affect drivers’ self-assessment and moreover, if drivers’ self-assessment can be improved. In the case of the Swedish driver education system, it might be possible to relate the self-assessments of knowledge and abilities to the theory test and the practical driving-license test. However, this needs to be further explored. Regarding the drivers’ self-assessment of motives for driving and personal goals, it seems more difficult to relate the self-assessment to an external measure. Therefore, one might need other strategies to judge if these goals are fulfilled.
Introduction

Content of driver education

During the last few years, the driver education in some countries (e.g., Sweden, Norway and Finland) has come to focus more on the drivers’ evaluation of their own competence instead of focusing only on their performance regarding driving-related knowledge and abilities (Hatakka, Keskinen, Gregersen, Glad, & Hernetkoski, 2002; Jonsson, Sundström, & Henriksson, 2003; Statensvegvesen, 2002). The driver education system in Norway states that it is important that the driver is aware of his or her knowledge, abilities, driver behaviour and goals. In Finland the test-takers evaluate their competence before they take the practical test. After the test, the driver’s self-evaluation is compared to the judgement made by the examiner. However, the result from the self-evaluation has no impact on the result of the practical driving test. The emphasis on self-evaluation is also evident in a new curriculum for driver education in Sweden. The new curriculum (VVFS2004:110) will be introduced in March 2006. Compared to the present curriculum, the new curriculum contains new parts focusing on the driver’s evaluation of his or her knowledge, ability and goals.

The new Swedish curriculum is based on the “Goals for Driver Education” framework, in short GDE framework (Hatakka et al., 2002). The GDE framework is a conceptual model of driver training and education. Similar to the model, the curriculum comprises two dimensions (see table 1). The first dimension consists of four hierarchical levels of driver behaviour. The first two levels, vehicle manoeuvring and mastery of traffic situations, are basic abilities which are included in the present driver education. The two following levels are about goals and context of driving as well as goals for life and skills for living. These goals are considered to be of great importance for how the driver decides to behave in traffic, and thus these goals are important for traffic safety (Hatakka et al., 2002). The second dimension in the GDE framework is formed by three goals for training: basic skills and knowledge, knowledge and skills concerning risk increasing factors and skills for self-evaluation. In the curriculum, the second dimension only comprises two parts, i.e. basic skills and knowledge and skills for self-evaluation. The level of knowledge and skills concerning risk increasing factors has been included in the two other levels.
Table 1. A model of the new curriculum for driver education in Sweden.

<table>
<thead>
<tr>
<th>Contents</th>
<th>Knowledge and skills</th>
<th>Self-evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goals for life and skills for living</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goals and context of driving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mastery of traffic situations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vehicle manoeuvring</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the present Swedish driver education system, the driver education is voluntary, except for six-hour risk training on slippery surface. Because of this, the driving-license test, which consists of a theory test and a practical test, is the only way for the Swedish National Road Administration (SNRA) to assess if the test-taker has the knowledge and ability stated in the curriculum (Jonsson et al., 2003). Therefore it is of great importance that the tests are reliable in the sense that the test-taker’s result on the test is independent of the test-version administered or what driver examiner assessed the performance. It is also very important that the test is valid in the sense that it measures the goals of the curriculum. There has to be conformity between the goals of the driver education, the education itself and the driving-license tests in order to ensure that the test measures what it is intended to measure (Henriksson, Sundström, & Wiberg, 2004). Therefore, when a new curriculum is introduced, the driver education and the driving-license test needs to be altered so that there is correspondence between the content of the curriculum and the tests (Appelgren, 2004; Edvall, 2004).

**Purpose**

Since the new curriculum contains goals for the drivers’ self-evaluation, new strategies for assessment have to be employed in order to measure these goals in a reliable and valid manner. Because self-assessment is a fairly new area in the field of driver education and driver testing, there is very limited knowledge concerning the measurement of such goals. The overall purpose of this study was to examine how self-assessment in general has been measured in other fields. The purpose was also to investigate how self-assessments of driver competence have been measured. Moreover, if these measurement strategies could be applied to the cur-
riculum for driver education in order to measure the goals concerning drivers’ self-evaluation.

Method

The structure of the literature study

In order to investigate how self-assessment has been measured in other fields, a literature study was conducted. In addition, the literature study also includes studies about drivers’ self-assessment. In this chapter, the strategy for the literature search is presented, and moreover, the concept of self-assessment and other related concepts are discussed. The next chapter includes the results of the literature study and consists of two major parts. Firstly, studies about self-assessment in general are presented and the relationship between self-assessment and external measures are reviewed. Secondly, studies that focus on drivers’ self-assessment are reviewed. In the third chapter, i.e. the discussion, conclusions from the studies reviewed are presented. The conclusions from the studies regarding self-assessment in general are compared with the conclusions from studies focusing on drivers’ self-assessment. Based on the conclusions, suggestions for further studies are made as well as suggestions for strategies for measuring drivers’ self-assessment in the Swedish driver education system.

Literature search

In order to collect relevant literature for the review, library databases and the literature databases Academic Search Elite, ERIC and ScienceDirect were used. Academic Search Elite is a database that covers different areas of academic study and ERIC covers literature on education. The database ScienceDirect includes Elsevier journals and was chosen since it contains a number of journals about traffic safety and traffic psychology. The databases were searched for articles about self-assessment and a number of articles were collected from different fields, mostly from psychology and education. The articles found were also searched for relevant references.
The concept of self-assessment and other related concepts

In the articles reviewed, many different concepts are used as synonyms of self-assessment and other concepts related to self-assessment. The concept *metacognition* includes two different aspects: control, monitoring and regulation of cognitive processes and knowledge about cognition, i.e. metacognitive knowledge (Bransford, Brown, & Cocking, 1999).

*Metacognitive knowledge* refers to students’ knowledge about their own cognition and control of their own cognition. Metacognitive knowledge comprises three different parts. Firstly, it comprises strategic knowledge, which is students’ knowledge of general strategies for learning and thinking. Secondly, it includes knowledge about cognitive tasks, which also refers to knowledge of when and why to use different strategies. Thirdly, metacognitive knowledge comprises self-knowledge, which is defined as knowledge about the self in relation to both cognitive and motivational components of performance (Flavell, 1979).

*Self-knowledge*, as mentioned above, is included in metacognitive knowledge. Self-knowledge is based on an individual’s own self-awareness and knowledge base. Self-knowledge could be divided into two major parts: *knowledge of one’s general cognition* and *beliefs about motivation*. The first part concerns knowledge of one’s strengths and weaknesses in relation to cognition and learning. It also includes awareness of the different types of general strategies students are likely to rely on in different situations. In addition to knowledge of one’s general cognition, individuals have beliefs about their motivation. These concern students’ judgements of their ability to accomplish a specific task, beliefs about the reasons students have for pursuing a task and students’ personal interest in a task as well as their judgements of how important the task is to them (Andersson et al., 2001).

When an individual makes judgements and evaluations based on self-knowledge, a *self-assessment* is made (Shrauger & Osberg, 1981). Self-assessment refers to the involvement of learners in making judgements about their own learning, particularly about their achievements and the outcomes of their learning (Boud & Falchikov, 1989).
The literature of self-assessment has been classified under three headings: conceptual, practical qualitative and quantitative (Boud & Falchikov, 1989). The most important parts of the conceptual framework are the literature about the reflective practitioner, teacher self-monitoring, metacognition and learning-how-to-learn. The practical qualitative group deals with the processes involved in introducing and using self-assessment in different situations. There are also discussions about the role of teachers in facilitating self-assessment. The quantitative group focuses on studies of student self-ratings compared to the ratings of students by teachers. Measures of agreement between teachers and students are often calculated and analyses of over-or underrating are conducted (Boud & Falchikov, 1989). The articles included in this review fit the description of the quantitative group.

The type of student self-assessment referred to most frequently in the literature is a process which involves teacher-set criteria and where students themselves carry out the assessment and marking. Another form of student self-assessment is the case where a student assesses herself or himself on the basis of criteria which she or he has selected. However, it is frequently the case that teachers control the assessment process, sometimes assessed by assessment experts. Students’ assessments and criteria are usually taken seriously but considered to be additional to the assessment undertaken by the teacher rather than replacing it (Dochy & McDowell, 1997).


In this review the concept of self-assessment is used even if other concepts that can be seen as synonyms of self-assessment were used in the original article cited.
Results
This chapter contains three different parts. Firstly, internal and external methods for assessment are discussed. Self-assessment can be viewed as internal assessment and is discussed in relation to other forms of external assessment where a test or an external judgement is made in order to assess the test-taker. Secondly, studies about self-assessment in general are reviewed. The relationship between self-assessments and external measures are reviewed and moreover, factors that affect the accuracy of self-assessment are investigated. Thirdly, studies about self-assessment of driver competence are reviewed.

Internal and external assessment

In order to assess the goals of the new curriculum, there are several possible methods that could be used. The present driving-license test can be seen as an external instrument for assessing the test-takers’ competence, consisting of an expert judgement of the drivers’ practical abilities and a standardized theory test for assessing the test-takers’ theoretical knowledge. Both tests are criterion-referenced, i.e. the purpose of the tests is to assess if the goals of the curriculum are fulfilled (Henriksson et al., 2004). Regarding the assessment of the goals concerning drivers’ self-evaluation, one might be able to use both external and internal assessments. An external assessment would be to let an expert judge how realistic the drivers’ view of their own knowledge and abilities seem to be. An internal assessment would be to let the students assess their knowledge, abilities and personal goals through self-reports.

The dominant model for psychological assessment has been a professional assessor gathering information about the person being judged, using procedures like interviews, standardized tests, historical data or systematic observations. Self-assessment is an alternative approach where the persons are asked to make judgements and evaluations based on their own self-knowledge. Self-assessment has several advantages. For example, individuals possess extensive information, which can be used to draw inferences about themselves, much more information than an external evaluator is likely to collect. The people being evaluated may also be aware of changes in their motives and attitudes that affect their behaviour and these changes may not always be elicited by external observers (Shrauger & Osberg, 1981). Another advantage of self-assessment is that it may decrease the time-investment teachers otherwise need to make in
their assessment. In addition, using self-assessment assists the development of self-assessment skills and self-criticism (Dochy & McDowell, 1997). Other advantages of self-assessments through questionnaires are their apparent simplicity, versatility and low cost as a method of data gathering (Schaughnessy & Zechmeister, 1997).

The critique against the validity of self-assessments is mainly based on two ideas. The first is that people are unaware of some of their most important feelings and are therefore often not capable of assessing themselves accurately. The second idea is that the responses of the persons might be biased (Shrauger & Osberg, 1981). Regardless of how carefully survey data are collected and analyzed, the value of these data depends on the truthfulness of the respondents’ answers to the survey questions. How willing should we be to believe that the survey responses reflect people’s true thoughts, opinions, feelings and behaviour (Breakwell, Hammond, & Fife-Schaw, 1995)? When people judge themselves, they have a tendency to present themselves to others in a socially desirable way, and therefore the self-assessment might not be valid (Shrauger & Osberg, 1981). By its very nature, survey research involves reactive measurement. Respondents know not only that their responses are being recorded, but they may also suspect that their responses may serve as the basis for some social, political or commercial action. Hence the pressures are strong for people to respond as they think they “should” and not as they actually do believe. The term often used to describe these pressures is social desirability (Breakwell et al., 1995). It should come as no surprise that sensitive and socially undesirable behaviours are often misrepresented, if reported at all. Reports of involvement in illegal practices are likely to be prone to error (Schaughnessy & Zechmeister, 1997). Sometimes researchers can examine the accuracy of verbal reports directly by using archival data or unobtrusive measures. In other cases, researchers use indirect observation of respondents’ behaviour in order to assess the accuracy of verbal reports (Breakwell et al., 1995).
Studies of self-assessment

Recently, there has been an upsurge of interest in self-assessment, as the role of self-assessment both in learning generally, and in the development of professional competence, has been recognised. One of the characteristics of effective learners is that they have a realistic sense of their own strengths and weaknesses and that they can use knowledge of their own achievements to steer their studying in productive directions. Self-assessment is formative in that it contributes to the learning process and assists learners to direct their energies to areas for improvement, and it may also be summative, either in the sense of learners deciding that they have learned as much as they wished to in a given area, or, in formal institutional settings, it may contribute to the grades awarded to students (Boud & Falchikov, 1989).

The idea that self-assessment is an important aspect is present both in the psychological and pedagogical field. In a revised version of Bloom’s taxonomy (Andersson et al., 2001), the importance of metacognitive skills is stressed. Compared to Bloom’s taxonomy, which only included a cognitive domain, the revised taxonomy contains an additional dimension, i.e. a knowledge dimension. The dimensions are considered to be at least partially hierarchical. The highest level in the knowledge dimension is metacognitive knowledge. Self-knowledge is an important aspect of metacognitive knowledge and includes knowledge of one’s strengths and weaknesses in relation to cognition and learning. The accuracy of self-knowledge seems to be crucial for learning. Teachers should not try to boost students’ self-esteem by providing them with positive but inaccurate feedback about their academic strengths and weaknesses. It is much more important for students to have accurate perceptions and judgments of their knowledge base and expertise than to have inflated and inaccurate self-knowledge. If students are not aware of what they do not know, it is unlikely that they will make any effort to learn new material. It is important that teachers help students make accurate assessments of their self-knowledge and not attempt to inflate students’ academic self-esteem (Pintrich & Schunk, 1996).
Self-assessment as an instrument for prediction

Some studies have compared the prediction accuracy of self-assessment with the prediction accuracy for external measures, and reviews of the empirical literature have offered favourable evaluations of self-assessments. Shrauger and Osberg (1981) showed that in 33 out of 43 studies, self-assessments were as good as, if not better than, test scores, grades or external rater evaluations in predicting academic performance, outcomes of psychotherapeutic interventions, job performance and peer ratings. However, Shrauger and Osberg (1981) emphasise that few studies of self-assessment involved actual situations of selection or decision-making. Therefore, they argue that it is not clear that the self-assessments will predict behaviour if significant stakes are involved (e.g., selection or promotion).

Another review (Mabe & West, 1982) investigated how accurately self-assessments of ability predicted performance measures. The mean correlation between self-assessments and performance was $r = .29$. Moreover, the correlations were related to nine favourable measurement conditions. The result showed that four conditions accounted for most of the prediction accuracy: expectation of self-assessment validation, self-assessment instructions using social comparison, self-assessment experience and instructions of anonymity. It was found that for 76 percent of the correlations, three or fewer of the nine measurement conditions were present. Thus it should not be surprising that many of the validity correlations were low (Mabe & West, 1982). The finding that the validity of self-assessment is higher in studies where subjects expect their self-assessments to be validated against external criteria was explained by Mabe and West (1982). They argued that individuals are influenced by two motives when giving self-reports: the desire to present themselves accurately and the desire to present themselves favourably. Instructions regarding cross-checking may thus give more accurate and valid self-assessments.

Relationship between self-assessment and external measures

Other studies have compared self-assessments with external measures. In a review of self-, peer- and co-assessment in higher education (Dochy, Segers, & Sluijsmans, 1999), it was concluded that research reports positive findings concerning the use of self-assessment in educational prac-
tice. Students in higher education are well able to self-assess accurately and this ability improves with feedback and development over time. Moreover, students who engage in self-assessment while learning tend to obtain a higher percentage of correct responses during learning trials on a test than those who learn without self-assessment.

In Boud and Falchikov (1989) student self-assessment was also related to teacher marks. One purpose of the review was to investigate if students tend to overrate or underrate themselves compared to teacher marks. The results showed that in most studies a greater number of student marks agree than disagree with staff marks. However, in most studies there was a tendency towards over- or underrating. A second purpose was to examine if students of differing abilities tend to over- or underrate themselves compared to teachers. The general trend in the studies reviewed suggests that high-achieving students tend to be realistic and perhaps underestimate their performance, while low-achieving students tend to overestimate their achievements probably to a greater extent than the underestimation. A third purpose was to investigate if student in different kinds or levels of class tend to over- or underrate themselves compared to teacher ratings. The trend of the studies examined was that students in later years of courses and graduates have a tendency to either become more accurate in their ratings, or tend to underestimate their performance. A fourth purpose was to examine if students overrate or underrate themselves when self-marks were used for assessment purposes. In general, it was difficult to determine an answer to this question, as studies do not consistently specify whether the marks of students are formally recognised. However, in those cases where student marks appear to count, students tend to overrate themselves.

In a meta-analysis (Falchikov & Boud, 1989), studies that compared self- and teacher marks were examined. One conclusion from the study was that expertise within a particular field is more influential on the accuracy of self-assessment than is seniority or duration of enrolment. Moreover, it was emphasised that self-assessment may be regarded as a skill and, as such, needs to be developed. It has been suggested that good assessment practice, whether ratings be made by students or by teachers, should include training of assessors.
In the field of employee selection self-assessments of competence have been related to external measures. For example, Love and Hughes (1994) presents a number of studies where the relationship between self-assessment for employee selection and written test scores was examined. The studies indicate that the self-assessments are valid for relevant applicant characteristics (typing ability, typing speed, spelling ability and word meaning for clerical workers) when linked with actual standardized test performance in these same areas. In a study by Love and Hughes (1994) the relationship between self-assessment ratings of police sergeant candidates and a department-specific promotional examination, measuring similar attributes, was explored. The participants (n=73) were required to provide ratings of their ability in 13 performance dimensions. An external measure consisting of the same 13 performance dimensions was developed. The self-assessment instrument included 65 items and was completed before the written examination. There was a positive significant correlation between the self-assessments and the written examination (ranging from r=.21 to r=.34).

It has been shown that there is a correspondence between self-assessment and external ability measures for a wide range of abilities and knowledge (Ackerman, 1997). Self-assessment of math/spatial ability and verbal ability has been shown to correlate (r=.58 and r=.42) with test composites of the math/spatial and verbal abilities, respectively (Ackerman, Kanfer, & Goff, 1995). Self-assessments of knowledge have also been compared with external measures of 32 different domains. The result showed that self-assessments of knowledge in Math and Physical Sciences and in Technology had substantial correlations with external measures of Math, Spatial and Mechanical abilities (Rolfhus & Ackerman, 1996).

In a study by Ackerman, Beier and Bowen (2002), one purpose was to investigate the agreement between self-assessments of knowledge and external measures of knowledge. Self-report measures were obtained and compared against external measures for the 230 participants. The external measure was a knowledge test battery based on a series of tests provided by the College Board and College Level Experience Program. The tests represented four broad domains: Science, Civics, Humanities and Business. These domains were divided into 18 subscales.
The external measure was related to three different self-report scales. The first was a scale for self-assessments of ability. This scale consisted of five areas of broadly described abilities and knowledge: verbal, math, mechanical, clerical and self-management. The instructions directed the participants to rate themselves compared with other persons of the same age. The second self-report scale consisted of a parallel set of 18 self-report knowledge scales for the 18 objective knowledge scales. The persons were asked to rate how much they knew about a specific topic. The third scale was a self-assessment of four areas of competencies and aptitudes: verbal, math, spatial and science. The subjects were told to consider whether they had the skill or ability stated. Each item provided a specific reference to a particular skill or ability (Ackerman et al., 2002).

The results showed a broad agreement between self-assessed and objective knowledge. When the external measure was compared with self-assessments, it was shown that Science knowledge correlated positively and significantly with self-reported Math, Spatial and Mechanical abilities, while Humanities, Civics and Business correlated positively with self-assessed verbal abilities. The correlation between self-assessments of knowledge and external knowledge measures ranged from $r=-.07$ to $.68$. The largest correlations were found for the Science domains. The Business domain failed to show a significant correlation with self-ratings (Ackerman et al., 2002).

In a cross-national study (Shen & Pedulla, 2000), the relationship between students’ achievement and self-perceived academic performance was investigated. In the study, data from the Third International Mathematics and Science Study (TIMSS) was used comprising primary and middle school students in about forty countries. Firstly, the relationship between self-perceived competence and achievement was examined. The results indicated that there was a significant positive relationship for almost all countries. Secondly, the relationship between achievement and self-perceived competence was compared among the countries. The results showed that there was a significant negative relationship, indicating that in countries where students’ achievements were relatively high, the students tended to rate their competence lower than did students in countries performing less well. Conversely, in countries where students’ actual achievement was relatively low, the students tended to rate their competence higher than in countries performing better. Thirdly, the relationship between self-perceived competence and self-perception of
the level of difficulty of science and mathematics was investigated. The within-countries analysis showed that students with high self-perceived competence were more likely to perceive the subjects as being easy and vice versa. The between-countries analysis showed that countries where students tend to have high self-perceived competence are also likely to perceive the two subjects as being easy and vice versa. Finally, the relationship between students’ achievement and their perceived difficulty level of the two subjects was also investigated at the country level. The results suggested that students in countries with relatively high achievement usually perceive mathematics and science as being hard, while students in countries with relatively low achievement usually perceive these subjects as being easy. In summary, the results indicated that countries performing relatively poorly on TIMSS tests have students who have relatively high self-perceptions about doing well in mathematics and science and who tend to perceive the two subjects as being easy. Conversely, countries that perform relatively well on TIMSS tests have students who have relatively low self-perceptions of their competence and who tend to perceive mathematics and science as being hard. One possible explanation of this is that low-performing countries have low academic standards and high-performing countries have high academic standards.

Over- and underestimation of competence

As mentioned above, studies indicate that there is a relationship between self-assessments and external measures. Still, several studies show that some people either over- or underestimate their competence. There are various factors that may moderate or contribute to the tendency to misjudge one’s abilities. These include knowledge of the domain being tested, the difficulty of the domain, the specificity/ambiguity of the ability being evaluated, desirability of the trait being evaluated and ability of the test-taker (Ackerman et al., 2002).

Relative or absolute judgement

It has been shown that self-assessments are affected by the scale used to measure them. When people assess their competence in broadly defined areas and when a relative scale is used, people rate themselves as above average. When the questions are specific rather than general and when an absolute scale is used, the self-assessment are more realistic (Ackerman et al., 2002).
Task difficulty

It has also been suggested that individuals underestimate their abilities compared to others when the task is hard, and overestimate their abilities when the task is easy (Kruger, 1999). This hypothesis was investigated by asking 37 students to compare themselves with their peers along a series of domains of varying difficulty. There were four easy abilities included in the questionnaire: operating a computer mouse, driving a car, riding a bicycle and saving money. The four difficult abilities included were: programming a computer, juggling, playing chess and telling a really good joke. As predicted, the participants’ assessment of their skill compared to their peers was related to the ease or difficulty of the ability domain. The easier ability, the higher participants’ estimates of how they compare with their peers. The majority of the participants thought they were above average in all four of the easy ability domains. In the more challenging domains, participants believed they were worse than their peers. The conclusion was that the participants’ judgements of how they compare with their peers were based more on their own skills than on their assessments of the skills of the comparison group. As a result, they thought they were above average when considering easy ability domains but thought they were below average when considering more challenging skills.

Gender differences

The results from studies of differences in levels of perceived competence between male and female students are fairly consistent. Girls in general have consistently been found to underestimate their ability, while boys tend to overestimate their ability (Phillips & Zimmerman, 1990). This was also shown in Nyström (2004), where the self-perceived competence in mathematics of girls and boys was related to their performance on a mathematics test. Two percent of the girls and 16 percent of the boys overestimated their grade on the test. Moreover, 41 percent of the girls and 22 percent of the boys underestimated their grade on the test. This pattern was also found when students’ answers to the question “how good are you in mathematics” were related to test-results.
Student competence

Other studies have indicated that competent students make more realistic self-assessment than less competent students. A study (Longhurst & Norton, 1997) that compared students’ self-assessment of a specific coursework essay with tutors’ ratings showed that there was a significant positive correlation between student and tutor grades ($r=.43$). This indicated that students are reasonably accurate at grading their own essays. A more detailed analysis showed that 16 percent of students predicted their essay grade accurately, 41 percent were within one grade and 70 percent were within two grades of difference. Almost exactly the same percentage of students overestimated (43%) as underestimated (41%) their performance. When the students were compared regarding performance on the essay, the lower scoring students overestimated their performance and the higher scoring students underestimated their performance.

The difficulty of making accurate self-assessments was also shown in a study of students’ evaluation of their academic performance. Participants were 49 freshmen taking a required course in World Arts, at a large university in the USA. High-achieving students were expected to show better metacognition, defined as accurate final grade predictions, than low-achieving students. Moreover, prediction inaccuracy, defined as the difference between predicted and actual grades, was expected to be negatively related to final grades. The results supported the expected relationship between achievement and metacognition: both the predicted grades and the prediction inaccuracy correlated significantly with the final grades. It has also been shown that students with high Grade Point Average (GPA) make more accurate predictions of their grades in psychology than students with low GPA. Low-achieving students overestimated their grades more than high-achieving students. High-GPA students were more confident about their grade estimates than were low-GPA students. In addition, research has shown that both low- and medium-ability students overestimated virtually all of their grades with confidence (Hartman, 2001).

The validity of self-assessments in a military setting was evaluated by Fox and Dinur (1988). The possibility that instructional manipulation could improve the validity of self-assessment was also investigated. The participants were 357 males who were being screened for a military training course. On the final day of selection the participants completed peer
evaluation measures as well as self-assessment measures. The participants’ field commanders were also asked to complete evaluations of their recruits. The participants were randomly assigned to an experiment group and a control group. The experiment group, in contrast to the control group, was told that their self-assessments could be compared with peer and supervisor evaluations as well as to objective ratings. The self-assessments, peer assessments and commanding officers’ assessments included ratings of physical fitness, efficiency, motivation to graduate, self-confidence etc. When the self-assessments were related to commander and peer ratings most of the correlations were significant, regardless of experimental treatment. Yet, correlations for the experimental group ratings were consistently higher when compared with those of the control group. When the self-assessments were compared with the commanders’ ratings, it was shown that the participants strongly inflated their ratings. Self-ratings were significantly higher than those of the commanders in each of the five dimensions compared (Fox & Dinur, 1988).

It has also been stated that the overestimation of competence occurs, in part, because people who are unskilled in these domains fail to realize that they are incompetent (Kruger & Dunning, 1999). Across four studies Kruger and Dunning (1999) investigated the participants’ self-assessment of competence. In the first study 65 participants rated their ability to recognize what is funny. A series of jokes was presented to the participants, who were asked to rate the humour of each one. Their ratings were compared with the ratings of experts, namely professional comedians. Afterwards the participants compared their ability to recognize what is funny with that of the average student by providing a percentile ranking. The results showed that self-assessments of ability were significantly correlated with the expert ratings ($r=0.39$). In addition the results showed that the participants tended to overestimate their ability relative to their peers. Those who performed particularly poorly relative to their peers grossly overestimated their ability. Participants scoring in the bottom quartile on the humour-test overestimated their performance by 46 percentile points. Participants in the other quartiles did not overestimate their ability to the same degree and those in the top quartile actually underestimated their ability relative to their peers.

In the second study participants’ self-assessment of their logical reasoning skills was assessed and compared with their result on a logical reasoning test. Similarly to the results in the first study, this study also showed that
participants that performed poorly overestimated their skills the most compared to their peers. In the third study self-assessed grammar knowledge was investigated and compared to the performance on a test of grammar. As in the two first studies the participants overestimated their ability and performance relative to objective criteria and the participant scoring in the bottom quartile grossly overestimated their ability relative to their peers. In this study the participants’ self-assessments failed to correlate with their actual test scores.

**Improving self-assessment**

In the study by Kruger and Dunning (1999), interventions were also made in order to investigate if self-assessments could be improved. Firstly, the participants were allowed to observe other participants’ performance on the test in order to gain an insight into their own performance. The results showed that the incompetent individuals failed to gain an insight into their own incompetence by observing the behaviour of others. For the high-performing participants, their self-assessments did improve by observing other people’s performance. Secondly, participants scoring in the bottom quartile on the logic test were trained in logical reasoning in order to investigate if raising the competence would provide a more accurate self-assessment. The results showed that the participants improved their self-assessment after their logical reasoning skills were improved. In contrast participants in the bottom quartile that did not receive the training continued to hold the mistaken impression that their performance was just fine.

In addition to the study by Kruger and Dunning (1999), some other studies have also indicated that the ability to assess one's own performance could be improved. Research has shown that novice learners are likely to engage in metacognitive activities less often and less successfully than learners with more subject area knowledge. However, training studies have demonstrated that students can learn to use metacognitive strategies, which may lead to better performance (Hartman, 2001). In the field of medicine, self-assessment of operative performance has been related to expert judgements of operative performance. Ward et. al. (2003) investigated the accuracy of self-assessment for the performance of a laparoscopic operation. Further, they also investigated if interventions lead to an improvement in self-assessment ability. The self-ratings were compared with expert-ratings. In order to study if interventions led
to an improvement of the self-assessment, the operations were videotaped so that the students could view their operation afterwards. The students evaluated their performance at three intervals: immediately after the performance, after self-observation of their videotaped performance and after a review of four videotaped benchmark performances. The correlation between experts’ evaluations and residents’ initial self-assessments was moderate ($r=.50$). The correlation between experts’ assessment and self-assessment after self-observation of the videotaped performance showed a statistically significant increase to $r=.63$, suggesting that the opportunity to view one’s own performance improved the self-assessment accuracy. However, the correlation between experts’ assessments and students’ self-assessments after they had viewed the benchmark videotapes was not significantly different at $r=.66$. Thus, the results suggest that reviewing one’s operative performance on videotape may aid the development of more accurate self-representations of performance, by stimulating focused self-reflection once completion of the task no longer constitutes a distraction.

In their review, Shrauger and Osberg (1981) discussed how to facilitate self-assessment. When people assess themselves, it is important for them to know as specifically as possible what aspects of their behaviour are to be assessed and in what contexts. Self-assessments are poor predictors primarily when people either misinterpret the situation to which they are to respond or have to make general judgements without clear behavioural referents. Typically, specific questions provide for better self-assessments than do more general questions. It is also very important to consider the phrasing of the questions. The questions have to be formulated so that they correspond as closely as possible to the behaviour to be assessed. Another aspect to consider is people’s motivation to be candid in their self-assessments. When the domain being assessed is highly desirable for the person, such as jobs or admission to training, it is assumed that the accuracy of self-reports relative to other criteria would be low. However, the motivation to present oneself in a favourable light should make self-assessments invalid only when it is stronger than the desire to present an accurate self-assessment.
Summarized conclusions

The studies presented in this section show in general that there is a relationship between self-assessments and external measures. However, some people either overestimate or underestimate their performance. Studies indicate that the accuracy of self-assessments is affected by several different factors. For example, expertise in a specific field is related to more accurate self-assessments. Several studies show that high-achieving students make more realistic self-assessments or tend to underestimate their competence compared to low-achieving students who tend to overestimate their competence. Some studies have also shown that students tend to overrate themselves when student marks appear to count. Moreover, it has also been shown that the types of items and scales used affect the accuracy of self-assessments. When broad items and a relative scale are used, students tend to overestimate their competence. Specific items and an absolute scale yield more accurate self-assessments. The difficulty of the domain assessed also seems to affect the self-assessment. Students’ judgements of their competence compared to those of their peers are based more on their own skills than on the assessment of the skills of the comparison group. As a result, students rate themselves as above average when the task is easy and below average when the task is hard.

Even if many of the studies reviewed suggest that many people either over- or underestimate their performance, studies have shown that self-assessments can be improved. For example, studies indicate that self-assessments are more valid when students are told that self-assessments might be crosschecked with external measures. For example, raising the competence of the students through education has shown to yield more accurate self-assessments. Further, it has also been shown that observation of one’s own videotaped performance results in more accurate self-assessments. It has also been stated that self-assessment may be regarded as a skill and, as such, needs to be developed. In order to facilitate self-assessment it is also important that the behaviour to be assessed is clearly defined. Another important aspect to consider is the phrasing of the questions. Questions have to be formulated so that they correspond closely to the behaviour that is to be assessed. Another aspect to consider is people’s motivation to be candid in their self-assessment. It has been argued that the motivation to present oneself in a favourable light makes self-assessments invalid only when it is stronger than the desire to present an accurate self-assessment.
Finally, studies have shown that self-assessment seems to have a positive affect on learning. Students who engage in self-assessment while learning tend to score higher during learning trials on a test compared to those who learn without self-assessment.

Self-assessment of driver competence

The overestimation of ability discussed in the previous section also seems to occur for car-drivers assessing their abilities to drive. In many studies, it has been found that drivers in general rate themselves as more competent than the average driver. For example, in a study conducted in Sweden and the United States, subjects judged their own driving skills in relation to other drivers (Goszynska & Roslan, 1989). The results showed that a majority of subjects regarded themselves as more skilful than the average driver. In a study conducted in France (Delhomme, 1991), the results showed that sixty percent of the subjects rated themselves superior to the average driver. The study also showed that drivers in general believe they commit fewer offences than the average driver.

Studies have also indicated that the self-assessment of driver competence is age- and gender-related. Studies show that young drivers consider themselves to be better than other drivers and the pattern has also been shown to be most typical among young men (Finn & Bragg, 1986; Matthews & Moran, 1986). The conclusion drawn in these studies is that young drivers are poor at assessing their own ability and judging risks adequately. They underestimate the risks and overestimate their skills as drivers.

In a study by McCormick, Walkey and Green (1986), another approach was used to study the self-assessment of driver competence. In the study the subjects rated “me as a driver”, “an average driver” and “a very good driver” on eight different scales: foolish-wise, unpredictable-predictable, inconsiderate-considerate, dangerous-safe, tense-relaxed, worthless-valuable, irresponsible-responsible. The result showed that a majority of drivers would rate themselves above average on a number of important characteristics, but also showed that they rated themselves below “a very good driver”.

20
Self-assessment of drivers’ competence has also been explored by relating the competence to a number of criteria rather than comparing the driver’s own competence with that of the average driver. Hatakka (1998) studied drivers’ self-assessments of driver competence in different situations. The self-evaluations were also rather detailed. The results showed that men and women differ in their assessment of risks and evaluations of themselves as a driver already before having any experience of independent driving. Women typically have a cautious attitude to driving and lower confidence in their personal skills. Male drivers expect a lower level of risks connected with road conditions and have higher confidence in their personal skills, although they acknowledge the risks regarding their careless driving habits. One factor common to both men and women was that a lower evaluated risk level and higher confidence in oneself as a driver were found in the younger age group in both genders. The results also showed that a high level of evaluated skills correlated with a low level of worrying, external risks and risks connected with the driver himself or herself.

Lajunen and Summala (1995) studied skill and safety-motive dimensions in drivers’ self-assessments of their driving abilities. Two factors were assumed to explain drivers’ self-assessment: those related to skilled and fluent driving (skill-factor) and the other related to safe and anticipatory driving (the safety-motive factor). In order to study drivers’ self-assessments, a questionnaire named the DSI, was developed. The aim of the DSI was to measure the drivers’ abilities in different aspects of driving, i.e. drivers were asked to identify the weak and strong components in their driving. The results showed that the DSI consisted of two factors, one safety-motive factor and one skill factor. The result also indicated that driving experience predicts scores obtained on the skill scale and on the motive scale. Experienced drivers (more than 5000 km/year) gained higher scores on the skill scale than inexperienced drivers. Experienced drivers also got lower scores on the motive scale than inexperienced drivers. The results indicate that driving experience increases drivers’ skill orientation but decreases their respect for driving safety and traffic rules. Inexperienced drivers emphasize safety-motives more than experienced drivers. Results have also shown that men report higher skill scores than women and that safety-skills correlate negatively with accidents and penalties. In addition, safety-skills correlate negatively with scores measuring speed and perceptual-motor skills correlate positively with number of penalties, life-time mileage, speed and a positive attitude to driving (Lajunen, Corry, Summala, & Hartley, 1998).
Relationship between self-assessment and actual driving

In some studies attempts have been made to relate self-assessments of driver competence to actual driving. In a study by Eby et al. (2003) the self-assessments of older drivers was measured with the “Driving Decisions Workbook”. The purpose of the workbook was firstly, for drivers willing and able to assess their own driving abilities, to give feedback for making good driving decisions by increasing self-awareness and general knowledge. Secondly, the purpose was to increase general awareness of age-related decline in driving abilities for generating discussion with peers and families. The aim of the study was to investigate if these purposes of the workbook were fulfilled and in addition if problems identified by drivers in the workbook related to problems they had with actual driving. A sample of 99 licensed drivers aged 65 and above was used. The participants completed the Driving Decisions Workbook as well as a questionnaire developed to identify self-reported increases in self-awareness and general knowledge after completing the workbook. In order to assess their actual driving a road test was taken. The participants drove a 7-mile route with 28 structured manoeuvres at specific locations. For each manoeuvre an examiner judged the performance. The results showed that after completing the workbook about three fourths of the participants reported being more aware of changes that could affect driving. All respondents found the workbook at least a little useful and thought that it could help to facilitate family discussions. The workbook responses were also positively correlated with overall road scores. Significant correlations were also noted between the road test and a majority of workbook subsection responses. In conclusion, the study indicated that the workbook might be a useful instrument and an educational tool for the older driver. It may encourage an older driver to drive more safely and/or seek clinical assessment (Eby et al., 2003).

The purpose of a study by Groeger and Grande (1996) was to identify what psychological variables underlie people’s overly positive assessments of their own ability and to explore the relationships between the self-assessments and actual ability. Self-assessed driving ability was measured by the Driver Behaviour Questionnaire (DBQ). Over 300 drivers assessed their driving ability in comparison to that of a novice. The results showed that driving experience was related to a positive view of one’s own driving ability. In the study, the actual driving ability was measured by a driving instructor for 114 of the original subjects. The subjects’ DBQ responses were compared to the assessments of their actual driving
on a 15 mile route. It was found that self-assessments did not relate to actual ability. Instead there was a relationship between the comments made by the instructor during the drive and the subject’s self-assessment as measured earlier in the study. There was also a relationship between the self-assessed skill reported in the questionnaire and the drivers’ feelings of how well they had driven on the particular occasion of the test drive.

In a similar study, Rimmö (2000) compared responses in the DBQ with driving behaviour in a short test drive. It was found that performance during a test drive correlated only with inexperience errors and not with violations, inattention errors or mistakes. Since the DBQ measures aberrant driver behaviours committed relatively rarely, it is very questionable if these behaviours would occur during a short test drive. In addition, social desirability bias might distort driving in a test drive as well as the DBQ responses.

Gregersen and Bjurulf (1996) described a study by (Moe, 1986) where self-assessed ability was compared with an objective measure of ability. The results showed a high correspondence between the self-assessed and the objective measure. The drivers’ speed was measured in the study and was related to self-assessed skill. The result showed that high-speed drivers believed themselves to be significantly more skilled than low-speed drivers. This difference was found for young men, not for women or older men.

Social desirability in self-evaluation of driver competence

Like many other self-reports of behaviour, traffic behaviour questionnaires are easily biased by socially desirable responding. Since self-reports as a measure of driving behaviour are more open to socially desirable responding than unobtrusive measures, there is a need to explore the effect of socially desirable responding in self-reports. Socially desirable responding has been divided into two types; impression management and self-deception. Impression management refers to the deliberate tendency to give favourable self-descriptions to others. Self-deception is defined as a positively biased but subjectively honest self-description (Lajunen, Corry, Summala, & Hartley, 1997; Lajunen & Summala, 2003).
A questionnaire for measuring socially desirable responding (SDR) in the traffic context was developed by Lajunen et al. (1997). The questionnaire was named *The Driver Social Desirability Scale* and measured two factors: driver impression management and driver self-deception. The questionnaire was related to a scale measuring driver behaviour called Driver Behaviour Inventory (DBI). The results showed that driver impression management was negatively related to self-reported number of accidents, punishments, overtaking frequency, speeding and driving aggression and positively related to traffic rule compliance. The results also indicated that driver self-deception correlated positively with variables measuring sense of control in traffic. The conclusion was that the Driver Social Desirability Scale is appropriate for measuring SDR in a traffic context and it is recommended that some instrument for controlling SDR be always used when traffic behaviour is investigated solely through self-reports.

**Summarized conclusions**

In summary, most studies that have investigated self-assessments of driver competence are studies where the drivers relate their competence to that of the average drivers. In general, these studies show that drivers rate themselves as more competent than the average driver. This pattern is specifically common for young male drivers who overestimate their ability and underestimate risks. In other studies, the self-assessment of driver competence has been investigated by using an absolute rather than a relative scale. Drivers have assessed their competence in relation to a number of criteria instead of comparing their competence with that of the average driver. The results show that women in general have a more cautious attitude to driving and lower confidence in their skills than male drivers who expect a lower level of risk and have higher confidence in their own skill. Moreover, young drivers consider the risks smaller and have higher confidence in their ability than older drivers.

Few studies have related self-assessed competence to actual driving. The results from the studies conducted indicate that self-assessed driver competence correlates with some aspects of actual driving. Finally, when self-assessments of driver competence are made, it is important to have in mind that the answers could be biased by socially desirable responding. Questionnaires for measuring socially desirable responding have been developed and it has been suggested that such instruments be used when traffic behaviour is measured through self-reports.
Discussion

Since the new curriculum for driver education contains goals for drivers’ self-evaluation, new strategies for assessment have to be employed in order to measure these goals. The purpose of this review of literature was to examine how self-assessment in general is measured in different fields. Secondly, the purpose was to investigate how self-assessments of driver competence have been measured. In this section the conclusions of the results are presented and suggestions for further studies are presented. In addition, the conclusions are related to the goals of the new curriculum with a focus on how these goals could be measured in a reliable and valid manner.

The relationship between self-assessment and external measures

In conclusion the studies reviewed show that there is a moderate relationship between self-assessment and external measures. Studies indicate that the accuracy of self-assessments is affected by several different factors. Expertise in a specific field is related to more accurate self-assessments. Several studies indicate that high-achieving students make more realistic self-assessments of their competence compared to low-achieving students, who tend to overestimate their competence. In addition, some studies have shown that students tend to overrate themselves when student marks appear to count. Moreover, it has also been shown that the types of items and scales used affect the accuracy of self-assessments. In formulating questions for self-assessment, one could distinguish between absolute and relative judgements. Absolute judgements assess whether the individual will meet a set criterion level of performance or will engage in a specific behaviour. Relative judgements involve assessing one’s behaviour relative to some reference group (Shrauger & Osberg, 1981). Studies indicate that when broad items and a relative scale are used, students tend to overestimate their competence. Specific items and an absolute scale tend to give more accurate self-assessments. The difficulty of the domain assessed also seems to affect the self-assessment. Students’ judgements of their competence compared to their peers are based more on their own skills than on the assessment of the skills of the comparison group. As a result, students rate themselves as above average when the task is easy and below average when the task is hard.
Although many of the studies reviewed suggest that some people either over- or underestimate their performance, some studies have shown that self-assessments can be improved. For example, when students are told that self-assessments might be crosschecked with external measures, self-assessments tend to be more valid. Raising the competence of the students through education and observation of their own videotaped performance are two other approaches that have shown to yield more accurate self-assessments. It has also been pointed out that self-assessment may be regarded as a skill and therefore needs to be developed. In addition, self-assessment seems to have a positive affect on learning. Students who engage in self-assessment while learning tend to score higher during learning trials on a test compared to those who learn without self-assessment.

The focus in this literature study has been on the relationship between self-assessments and external measures and the accuracy of the self-assessments. However, Boud and Falchikov (1989) pointed out that self-assessment can have important implications even if the congruence between ratings is not maximal. Self-assessment can be a valuable learning activity, even in the absence of significant agreement between student and teacher, and can provide potent feedback to the student about both learning and educational and professional standards. It is important to develop ways in which systematic formative self-assessment activities can be incorporated into courses to improve student skills in making sensitive and aware judgements on their own work.

In this literature study, studies of self-assessments of driver competence were also included. In most of these studies the self-assessed competence of the driver is related to that of the average drivers. In general, these studies show that drivers rate themselves as more competent than the average driver. This pattern is specifically common for young men who overestimate their ability and underestimate risks. Other approaches have also been made where self-assessment has been investigated by using an absolute rather than a relative scale. In this case, drivers assessed their competence in relation to a number of criteria instead of comparing their competence with that of the average driver. Similarly to the other studies, the results show that females in general have a more cautious attitude to driving and lower confidence in their skills than male drivers who expect a lower level of risk and have higher confidence in their own skill. More-
over, young drivers consider the risks to be smaller and have higher confidence in their ability than older drivers.

In many studies, drivers have assessed their competence both in relation to other drivers and in relation to different criteria. However, the self-assessments have rarely been compared with external measures. A few studies have related self-assessed competence to actual driving. The results from these studies indicate that self-assessed driver competence correlates with some aspects of actual driving. In order to explore the validity of the drivers’ self-assessments, there is a need to explore this field further and relate self-assessments of knowledge and ability to external measures. When self-assessments of driver competence are made, it is also important to have in mind that the answers could be biased by socially desirable responding. Questionnaires for measuring socially desirable responding have been developed and it has been suggested that such instruments be used when traffic behaviour is measured through self-reports.

Assessing the goals of driver education

The new curriculum for driver education is planned to be introduced in March 2006. When the goals of the driver education are altered, it is important that the education and the tests are altered too, so that there is a correspondence between the three parts. As mentioned above, the curriculum includes goals where drivers should assess their own ability, knowledge, motives and goals. A reliable and valid measurement of such goals has been the main focus of this report. Boud (1990) stressed that there is often a gap between the issues emphasised in education and the assessment practices used to measure what the students have learned. There is too often a discrepancy between the high level of course objectives and assessment tasks. Therefore, there is a need to examine the assessment practices to see if they are compatible with the goals for higher education. It has been stated that assessment encourages students to focus on those topics which are assessed at the expense of those which are not. In other words, assessment tasks define the instruction. This could imply that students are learning for the test rather than trying to internalise and make sense of the subject (Boud, 1990). In the same way, sound assessment practices can be used to improve instruction. Appropriately used educational assessments are potent tools that enhance the instructional process. An important condition for appropriate test use is that
tests are linked or integrated both with instructional materials and procedures (Dochy & McDowell, 1997). It has also been stressed that an important concept that links assessment with the quality of learning is that of consequential validity (Dochy et al., 1999). According to Messick (1989) it is important to consider the social consequences when judging the validity of the assessment. Judging validity in terms of whether a test serves its intended purpose requires evaluation of the intended or unintended social consequences of test interpretation and test use. If adverse social consequences are empiricallytraceable to sources of test invalidity, then the validity of test use is jeopardized.

Regarding the goals of self-assessment in the curriculum for driver education, the nature and content of the goals affects the strategy for measurement of self-assessment. If the goals refer to the student’s knowledge or abilities, a self-assessment could be made. However, since the method of self-reports is prone to problems like social desirability, and therefore not always as reliable as desired, the self-assessment could be related to an external measure in order to evaluate the agreement between the two measures. If the goals of self-assessment refer to the student’s knowledge, the self-assessment can be compared with an objective knowledge test. If the goals concern the practical driving abilities, the self-assessment of ability could be compared with an expert’s rating. If there is agreement between the self-assessment and the external measure, the test-taker would be considered to have a realistic view of his or her competence, and thus the goals of the curriculum are fulfilled. In the case where there is a discrepancy between self-assessment and external measures, the accuracy of the self-assessment needs to be improved with different interventions. In this case education might be used to raise the competence of the test-taker and hopefully provide more accurate self-assessments. Moreover, the self-assessment could be carried out at different times. One possibility is to conduct the self-assessment before the external measure and another possibility is to administer the self-assessment shortly after the external measure. Which method is to be used is dependent on which of them gives the most valid measurement of drivers’ self-assessment.

Some goals in the curriculum concern goals for driving and personal goals that affect driving. It might be more difficult to relate the self-assessment of these goals to an external measure and therefore more difficult to judge the validity of the self-assessment for these goals. In these cases, one possibility might be to measure the self-assessment of higher hierarchical levels through lower levels. It might be the case that test-
takers with an inaccurate self-assessment have an unrealistic view of their competence at all hierarchical levels. If so, a discrepancy between self-assessment and external measures at the lower levels would indicate that there is a discrepancy at the higher levels too. Another strategy for measuring self-assessment of the higher hierarchical levels would be to observe the student for a period of time, in order to make sure that these goals are fulfilled. These observations are most easily obtained in the driver education. However, if the goals of drivers’ self-evaluation are to be measured in the education, it is important to have a strategy for this education. What matters should be included in the education and what strategies for learning should be used? In a study by Eby et al (2003), a workbook was used in order to give older drivers feedback for making good driving decisions by increasing self-awareness and general knowledge. The study indicated that the workbook helped to improve drivers’ self-awareness and encouraged drivers to drive more safely. In addition, the workbook seemed to be a valid instrument since its responses corresponded to actual driving. It is possible that this approach using a workbook might be a successful instrument and educational tool even for younger drivers.

Another important aspect when using education to assess if the goals are fulfilled is that the driver education is of high quality and that all students participate in the education. One way to ensure that all students take part in the education is to introduce a compulsory course that will cover the goals. Another very important aspect is the competence of the teachers, since in this case they are the instruments for assessing if the goals are attained. The instructors must have sufficient competence and pedagogical skill so that they are able to judge if the students have acquired the goals stated in the curriculum.

The importance of educating teachers when a new curriculum is introduced can be exemplified by the change of the Swedish curriculum for the compulsory school system. A criterion-referenced curriculum, Lpo 94 (Skolverket, 1988), was introduced in 1994. Compared to the old norm-referenced grading system, the new system is criterion-referenced (Skolverket, 2004a). An examination of the grading procedure showed that many teachers had insufficient knowledge of curriculum and grading criteria, which can result in great differences in the grading. When the curriculum was introduced, there was also a lack of discussions related to grading, what matters to be assessed and if the matters assessed
correspond to the objectives in the curriculum (Skolverket, 2004b). When a new curriculum for driver education is introduced, it is therefore important that this mistake is not repeated. The driving instructors need to be educated and informed so that their instructions correspond to the goals in the curriculum and so that they can make a proper assessment of whether the students have fulfilled the goals in the curriculum or not.

Future research

The conclusions of the studies of self-assessment in general show a rather homogenous picture. There seems to be a relationship between self-assessment and external measures. The studies also show that a number of factors affect the accuracy of the self-assessment and that the self-assessment can be improved by different means. The studies of drivers' self-assessment also support these conclusions to some degree. However, self-assessment of driver competence has rarely been related to an external measure, and hence there is little knowledge of the relationship between self-assessed driver competence and external measures. There is also a need to know if the factors that seem to affect the accuracy of self-assessment in general also affect self-assessment of driver competence. For example, Kruger and Dunning (1999) showed that the type of judgement used (relative or absolute) might affect the accuracy of self-assessment. Since many of the self-assessments of driver competence are based on a relative judgement, it would be interesting to investigate if an absolute scale would result in more accurate self-assessments when compared to an external measure. Moreover, it has also been shown that the difficulty of the domain affects the accuracy of the assessment (Kruger, 1999) and that people tend to overrate themselves in comparison to others when the task is easy. Car-driving might be viewed as an easy task by many people and therefore it might be the case that people over-estimate their driver competence when they compare themselves with other drivers. Moreover, there is also a need to investigate if self-assessment of driver competence can be improved by different interventions, for example education or observation of performance.

The focus of this report has been on self-assessment of knowledge and abilities and on how to measure this via tests. Some goals of the curriculum for driver education might not be possible to measure via tests and therefore one might need other strategies for assessment. However, for other goals a self-assessment could probably be made and this self-
assessment could also probably be related to an external measure in order to check the validity of the self-assessment. When developing an instrument for self-assessment of driver competence, it is important to consider the factors contributing to assessment accuracy that were mentioned above. For example, are drivers more likely to overestimate their competence if marks count? Are drivers more prone to overestimate their ability if a relative judgement is made instead of an absolute judgement? Moreover, can the accuracy of self-assessment of driver competence be improved through observation of one’s own performance, observation of peer-performance or by raising the competence through education? Further, does a comparison with an external criterion make the self-assessments more accurate? These questions are all of interest for the development of a valid and reliable instrument for measuring self-assessment of driver competence and should preferably be investigated in further studies.
References


EDUCATIONAL MEASUREMENT

Reports already published in the series

EM No 1. SELECTION TO HIGHER EDUCATION IN SWEDEN. Ingemar Wedman

EM No 2. PREDICTION OF ACADEMIC SUCCESS IN A PERSPECTIVE OF CRITERION-RELATED AND CONSTRUCT VALIDITY. Widar Henriksson, Ingemar Wedman

EM No 3. ITEM BIAS WITH RESPECT TO GENDER INTERPRETED IN THE LIGHT OF PROBLEM-SOLVING STRATEGIES. Anita Wester

EM No 4. AVERAGE SCHOOL MARKS AND RESULTS ON THE SWESAT. Christina Stage

EM No 5. THE PROBLEM OF REPEATED TEST TAKING AND THE SweSAT. Widar Henriksson

EM No 6. COACHING FOR COMPLEX ITEM FORMATS IN THE SweSAT. Widar Henriksson

EM No 7. GENDER DIFFERENCES ON THE SweSAT. A Review of Studies since 1975. Christina Stage

EM No 8. EFFECTS OF REPEATED TEST TAKING ON THE SWEDISH SCHOLASTIC APTITUDE TEST (SweSAT). Widar Henriksson, Ingemar Wedman

1994

EM No 9. NOTES FROM THE FIRST INTERNATIONAL SweSAT CONFERENCE. May 23 - 25, 1993. Ingemar Wedman, Christina Stage

EM No 10. NOTES FROM THE SECOND INTERNATIONAL SweSAT CONFERENCE. New Orleans, April 2, 1994. Widar Henriksson, Sten Henrysson, Christina Stage, Ingemar Wedman and Anita Wester

EM No 11. USE OF ASSESSMENT OUTCOMES IN SELECTING CANDIDATES FOR SECONDARY AND TERTIARY EDUCATION: A COMPARISON. Christina Stage

EM No 12. GENDER DIFFERENCES IN TESTING. DIF analyses using the Mantel-Haenszel technique on three subtests in the Swedish SAT. Anita Wester

1995

EM No 13. REPEATED TEST TAKING AND THE SweSAT. Widar Henriksson
<table>
<thead>
<tr>
<th>EM No</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>AMBITIONS AND ATTITUDES TOWARD STUDIES AND STUDY RESULTS. Interviews with students of the Business Administration study program in Umeå, Sweden. Anita Wester</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>EXPERIENCES WITH THE SWEDISH SCHOLASTIC APTITUDE TEST.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>16</td>
<td>NOTES FROM THE THIRD INTERNATIONAL SweSAT CONFERENCE.</td>
<td>Christina Stage, Widar Henriksson</td>
</tr>
<tr>
<td>17</td>
<td>THE COMPLEXITY OF DATA SUFFICIENCY ITEMS.</td>
<td>Widar Henriksson</td>
</tr>
<tr>
<td>18</td>
<td>STUDY SUCCESS IN HIGHER EDUCATION. A comparison of students admitted on the basis of GPA and SweSAT-scores with and without credits for work experience. Widar Henriksson, Simon Wolming</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1996</strong></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>AN ATTEMPT TO FIT IRT MODELS TO THE DS SUBTEST IN THE SweSAT.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>20</td>
<td>NOTES FROM THE FOURTH INTERNATIONAL SweSAT CONFERENCE.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td></td>
<td><strong>1997</strong></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>THE APPLICABILITY OF ITEM RESPONSE MODELS TO THE SweSAT.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>22</td>
<td>ITEM FORMAT AND GENDER DIFFERENCES IN MATHEMATICS AND SCIENCE.</td>
<td>Anita Wester, Widar Henriksson</td>
</tr>
<tr>
<td>23</td>
<td>DO MALES AND FEMALES WITH IDENTICAL TEST SCORES SOLVE TEST ITEMS IN THE SAME WAY?</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>24</td>
<td>THE APPLICABILITY OF ITEM RESPONSE MODELS TO THE SweSAT.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>25</td>
<td>THE APPLICABILITY OF ITEM RESPONSE MODELS TO THE SweSAT.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>26</td>
<td>THE APPLICABILITY OF ITEM RESPONSE MODELS TO THE SweSAT.</td>
<td>Christina Stage</td>
</tr>
<tr>
<td>27</td>
<td>DIFFERENTIAL ITEM FUNCTIONING (DIF) IN RELATION TO ITEM CONTENT.</td>
<td>Anita Wester</td>
</tr>
</tbody>
</table>

1998

EM No 29. A Comparison Between Item Analysis Based on Item Response Theory and on Classical Test Theory. A Study of the SweSAT Subtest WORD. Christina Stage

EM No 30. A Comparison Between Item Analysis Based on Item Response Theory and on Classical Test Theory. A Study of the SweSAT Subtest ERC. Christina Stage

EM No 31. NOTES FROM THE SIXTH INTERNATIONAL SWESAT CONFERENCE. San Diego, April 12, 1998. Christina Stage

1999

EM No 32. NONEQUIVALENT GROUPS IRT OBSERVED SCORE EQUATING. Its Applicability and Appropriateness for the Swedish Scholastic Aptitude Test. Wilco H.M. Emons

EM No 33. A Comparison Between Item Analysis Based on Item Response Theory and on Classical Test Theory. A Study of the SweSAT Subtest READ. Christina Stage

EM No 34. Predicting Gender Differences in WORD Items. A Comparison of Item Response Theory and Classical Test Theory. Christina Stage

EM No 35. NOTES FROM THE SEVENTH INTERNATIONAL SWESAT CONFERENCE. Umeå, June 3–5, 1999. Christina Stage

2000

EM No 36. TRENDS IN ASSESSMENT. Notes from the First International SweMaS Symposium Umeå, May 17, 2000. Jan-Olof Lindström (Ed)

EM No 37. NOTES FROM THE EIGHTH INTERNATIONAL SWESAT CONFERENCE. New Orleans, April 7, 2000. Christina Stage

2001


EM No 39. PERFORMANCE AND AUTHENTIC ASSESSMENT, REALISTIC AND REAL LIFE TASKS: A CONCEPTUAL ANALYSIS OF THE LITERATURE. Torulf Palm
EM No 40. NOTES FROM THE NINTH INTERNATIONAL SWESAT CONFERENCE. Umeå, June 4–6, 2001, Christina Stage

2002

EM No 41. THE EFFECTS OF REPEATED TEST TAKING IN RELATION TO THE TEST TAKER AND THE RULES FOR SELECTION TO HIGHER EDUCATION IN SWEDEN. Widar Henriksson, Birgitta Törnkvist

2003

EM No 42. CLASSICAL TEST THEORY OR ITEM RESPONSE THEORY: THE SWEDISH EXPERIENCE. Christina Stage

EM No 43. THE SWEDISH NATIONAL COURSE TESTS IN MATHEMATICS, Jan-Olof Lindström

EM No 44. CURRICULUM, DRIVER EDUCATION AND DRIVER TESTING. A comparative study of the driver education systems in some European countries. Henrik Jonsson, Anna Sundström, Widar Henriksson

2004

EM No 45. THE SWEDISH DRIVING-LICENSE TEST. A Summary of Studies from the Department of Educational Measurement, Umeå University. Widar Henriksson, Anna Sundström, Marie Wiberg

EM No 46. SweSAT REPEAT. Birgitta Törnkvist, Widar Henriksson

EM No 47. REPEATED TEST TAKING. Differences between social groups. Birgitta Törnkvist, Widar Henriksson

EM No 49. THE SWEDISH SCHOLASTIC ASSESSMENT TEST (SweSAT). Development, Results and Experiences. Christina Stage, Gunilla Ögren

EM No 50. CLASSICAL TEST THEORY VS. ITEM RESPONSE THEORY. An evaluation of the theory test in the Swedish driving-license test. Marie Wiberg

EM No 51. ENTRANCE TO HIGHER EDUCATION IN SWEDEN. Christina Stage

Em No 52. NOTES FROM THE TENTH INTERNATIONAL SWESAT CONFERENCE. Umeå, June 1–3, 2004, Christina Stage

2005