



Agreement between driver examiners' assessments in a system using competency-based assessments

**Anna Wänglund
Umeå University, Sweden**



The Swedish driving-license system

Criterion-referenced system

- Curriculum
- Driver education
 - Mandatory
 - Voluntary
- Driving-license test
 - Theory test (computerized)
 - Practical driving test



Driving test

- Competency-based
- Overall assessment of candidate's competence with respect to four areas:
 - Vehicle knowledge/manoeuvring
 - Environmentally-friendly driving
 - Traffic regulations
 - Traffic safety/behaviour
- Flexible testing-routes
- Content of the test should meet objectives
- Guidelines for test content



Quality issues in driver testing

- We use test result to make the decision of issuing a driving license.
- We want the test to lead to desirable consequences – important that the test is of high quality
- The test has to measure what we want to measure in a **reliable** way
 - Test-retest reliability
 - Inter-rater reliability
- The test result has to be a **valid** indicator of the construct we want to measure.
 - Sample the content of the curriculum
 - Adequately measure driving skills important for safe driving
 - Predict safe drivers
- If a test lacks reliability it cannot be valid – therefore it is important to examine reliability.



Aim of the study

- To examine the inter-rater reliability of the Swedish driving test.
 - Are there differences between the assessments of two driving examiners assessing the same candidate?
 - Can these differences be explained by variables related to the examiners, to the candidate or to the test situation?



Data collection

- The study was conducted from October 2008 to January 2009.
- The ordinary and supervising examiner assessed the same candidate (n =535). 86 examiners and 5 supervising examiners.
- Examiners were selected to be representative of the situation that the majority of the candidates face – examiners having conducted 700 tests or more during 2006 were selected.
- Each examiner was accompanied by a supervising examiner during 1 day (assessing 7 driving tests).
- Both examiners assessed candidate performance. Supervising examiner also assessed the ordinary examiner's performance on the test.
- The sample can be considered representative for the population.



Examiner assessment

- Assessment of the candidate's overall performance
 - Pass/fail
- Ratings of:
 - How difficult it was to make the pass/fail decision
 - When, during the test, the pass/fail decision was made
- In cases where the examiners' assessments differed - possible explanations for differences

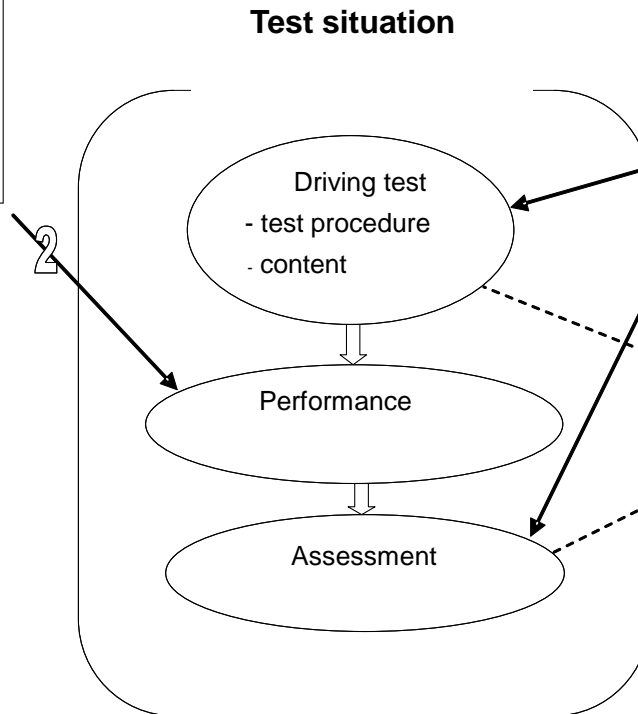


Agreement pass/fail assessment

Supervising examiner

		Fail	Pass	Total
Ordinary examiner	Fail	296	14	310
	Pass	23	202	225
	Total	319	216	535

Test taker
Sex, age
Registration category
Driving lessons
Self-assessment, attitudes to the test
Test anxiety



Ordinary driver examiner
Sex
Age
Education
Professional experience

Supervising driver examiner
Sex
Age
Education
Professional experience



Differences between examiners

In those cases where the examiners made different assessments in terms of pass/fail, there were differences regarding:

- How difficult it was to make the decision of pass/fail
- When the pass/fail decision was made during the test
- Whether the assessment was a holistic assessment



Discussion and conclusions

- There was a high agreement between examiners. When two examiners assess the same test they agree on the assessment in terms of pass/fail in 93 per cent of the cases.
- In those cases where there were differences between examiners there were only a few systematic differences with respect to candidate or examiner variables. These are not problematic with respect to consistency of assessment.
 - Difficulty of pass/fail decision
 - When pass/fail decision was made
 - Holistic assessment
 - Examiners perceived situation differently due to location in the car



Discussion and conclusions, cont'd

- The level of examiner agreement is good from an international perspective.
 - Norwegian study: examiners agree in 72 % of tests
 - British study of test-retest reliability: 64 % same result both tests
- High examiner agreement is an indicator of test quality. However, test quality is also affected by other factors.
 - Content of test routes
 - Difficulty of test routes
 - Differences between examiners within, between driving test centres?



Thank you for your attention!



Assessment – 6 point scale

1. Fail. Very poor driving, severe lack of competence in several areas.
2. Fail. Poor driving, lack of competence in several areas.
3. Fail. Lack of competence in one isolated area that are of significance for traffic safety.
4. Pass. Lack of competence not particularly significant with respect to traffic safety.
5. Pass. Good driving performance.
6. Pass. Very good driving performance.

Supervising examiner assessment

		1	2	3	4	5	6	Total
Ordinary examiner assessment	1	16	15					31
	2	21	157	24	7			209
	3		39	23	6	1		69
	4	1	12	8	102	22		145
	5		2		29	38	2	71
	6				2	5		7
Total		38	225	55	146	66	2	532



Differences on the 6-point scale

Per cent of tests

Supervising examiner higher assessment	14.3
Same assessment	63.3
Ordinary examiner higher assessment	22.4