Misclassifications in Preparatory VET

RCEC Research Center voor Examinering en Certificering

K-H Examens

CITO now you know

RCEC Workshop, 13 October 2010
Marianne Hubregtse, Theo Eggen, Huub Verstralen

K-H

Research questions

- How many 'wrong decisions' do we make

How is the % misclassifications influenced by:

1. The underlying **distribution** of simulated θ
2. The **standard** for separate exams
3. The **decision rule** for the diploma decision

K-H

Set-up of the study

- **Simulation** based on empirical data from Preparatory VET
  - Divided in 4 'exams' or competences:
  - Creativity (7 items)
  - Knowledge (8 items)
  - Social Skills (10 items)
  - Tasks (8 items)

  } 1 diploma

    - discarded items (9 items)
  - N=473
  - Students were assessed on all items

K-H

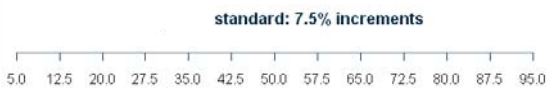
Influence of distribution (RQ 1)

Compare:

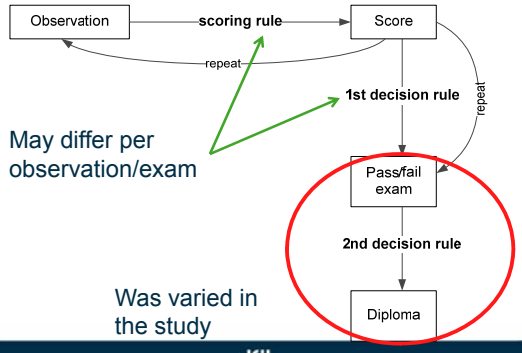
1. **Empirical** multivariate distribution
2. **Normalized** multivariate empirical distribution
3. **Standard normal** multivariate empirical distribution (with average correlation)

Influence of standards (RQ 2)

- Vary standard (pass/fail cut-off) of separate exams
  - Based on percentage of students passing
  - 5% through 95% pass percentage:



Influence of decision rules (RQ 3)



May differ per observation/exam

Was varied in the study

Different decision rules

Types of decision rule used: (and their names)

- Conjunctive: **no compensation**
  - All exams should be passed
- Compensatory: **1 or 2 compensations**
  - There are possibilities for compensation between exams
- Completely compensatory: **mean**
  - The exams are passed on average

Divide exam into 'grades'

- To determine whether one can compensate

	lowest theta value	
0 no compensation		f
	lowest cut-off theta value	a
1 lowest category		i
	lower cut-off theta value	l
2 low category		
	cut-off theta value	
3 pass		
	higher cut-off theta value	p
4 high category		a
	highest cut-off theta value	s
5 highest category		s
	highest theta value	

1 compensation possible

- Compensate a 2 with either a 4 or 5

	lowest theta value	
0 lowest category		f
	lowest cut-off theta value	a
1 lower category		i
	lower cut-off theta value	l
2 low category		
	cut-off theta value	
3 pass		
	higher cut-off theta value	p
4 high category		a
	highest cut-off theta value	s
5 highest category		s
	highest theta value	

2 compensations possible

- Compensate a 1 with a 5, a 2 with either a 4 or 5

	lowest theta value	
0 lowest category		
	lowest cut-off theta value	f
1 lower category		a
	lower cut-off theta value	i
2 low category		l
	cut-off theta value	
3 pass		
	higher cut-off theta value	p
4 high category		a
	highest cut-off theta value	s
5 highest category		s
	highest theta value	

Mean should be 'pass'

- Compensate anything

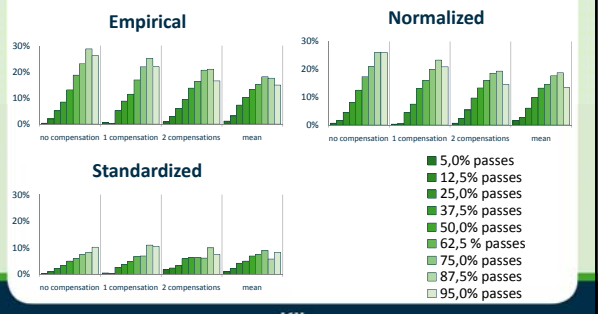
	lowest theta value	
0 lowest category		
	lowest cut-off theta value	f
1 lower category		a
	lower cut-off theta value	i
2 low category		l
	cut-off theta value	
3 pass		
	higher cut-off theta value	p
4 high category		a
	highest cut-off theta value	s
5 highest category		s
	highest theta value	

Measure of influence

- % misclassification = undue failure + undue pass

		Diploma decision	
		Pass	Fail
True ability	Competent	Due pass	Undue failure
	Incompetent	Undue pass	Due failure

Standard normal distribution yields lowest % of misclassification.



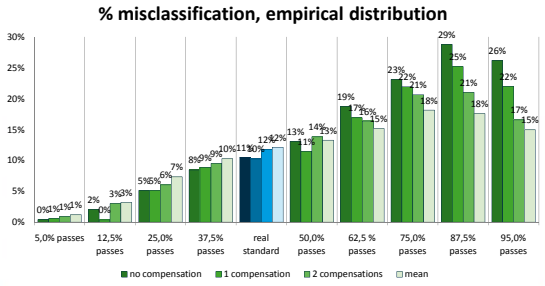
K+H

However…

- In practice, we usually have to deal with the empirical distribution.
  - We cannot change the students…
- We can change the standards!
- So, what is the influence of standards on the % misclassification?

K+H

Extreme standards minimize % of misclassification



K+H

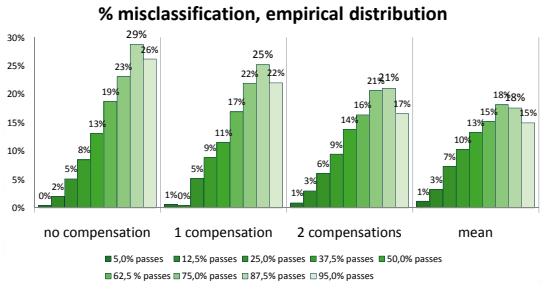
However…

- Extreme is usually 0% passing (or failing) the exam
- That is **not realistic**

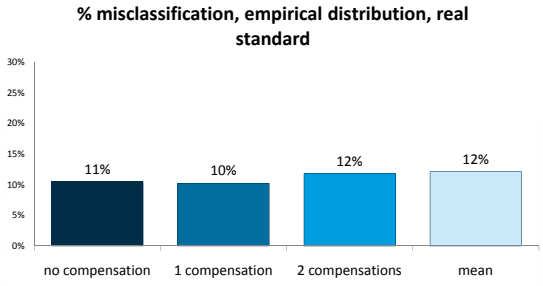
- More realistic is the real standard, which lies usually around 60% of the items correct
  - it still depends on the population how many pass

- Question remaining: what is the influence of the decision rule to be used?

Completely compensatory decision rules lower the % misclassification,



… but the effect is not visible at the real standard.



Recommendations

- Use completely compensatory decision rules
  - In essence this lengthens your exam, making the decision more reliable
- We cannot change the population; we should fit the exam to the population in design
- Adapt the standard of the exam to the individual
  - adaptive testing gives high information around the individual ability, and lessens measurement error

Thank you!

e-mail for more information or the presentation:
m.hubregtse@kch.nl

RCEC Research Center voor Examinering en Certificering
KH Examens
CITO now you know

weten waar je staat

Websites

Please find more information on:

- http://www.rcec.nl/
- http://www.kchandel.nl/