



Marking to the Common Marking Scheme with Criteria & Decision Rules

Paul Anderson
<dcspaul@ed.ac.uk>

<http://homepages.inf.ed.ac.uk/dcspaul>

Marking with criteria & decision rules

Motivation

- Returned marks which don't correspond to the CMS criteria
 - is everyone really "exceptional" ?
- Tenuous/opaque link between marks and learning outcomes
- Student dissatisfaction and lack of understanding of the marks/feedback
 - why did she pass (with 56%) and I failed (with 54%) ?
- Marker variability

My research

- PMark software for evaluation of criteria-based decision rules
- Evaluation (across several schools)

This talk

- A bit of background on the CMS & criteria-based marking
- What we learned about designing criteria-based marking schemes
- A very brief overview of PMark



Background

The Common Marking Scheme

The CMS

- Specifies a number of grades, together with a corresponding mark range and description of the requirements for each grade. For example:

50-59 C (Good): *The work clearly meets requirements for demonstrating the relevant learning outcomes.*

80-89 A2 (Excellent): *Outstanding in some respects, the work is often beyond what is expected of a competent student at their level of study. Demonstrates that the student is actively extending their knowledge and capacity well beyond required materials and making new connections independently.*

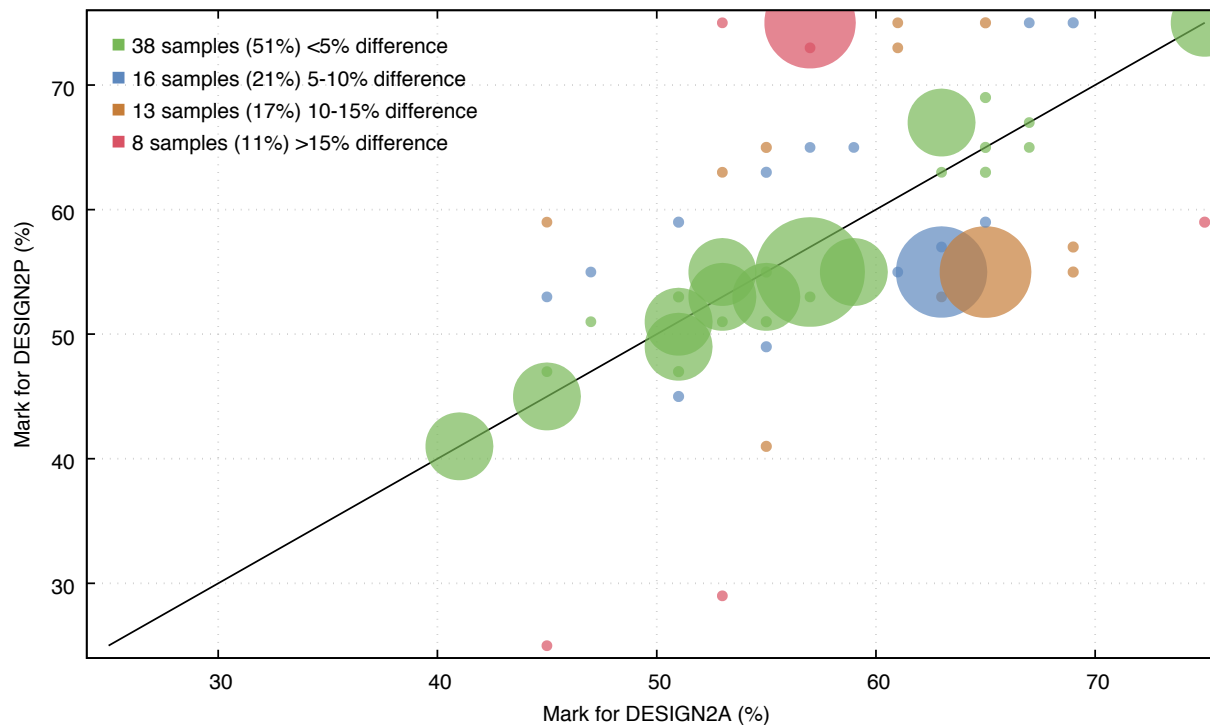
I think ...

- The CMS grades are a good basis for assessing "divergent" tasks
- In most cases, it is not realistic to expect the fine-grained percentage mark to be reliable (but we are forced to generate them)

https://web.inf.ed.ac.uk/sites/default/files/atoms/files/informatics_extended_cms-2019-09-16.pdf

Variability

"we mislead students that there is something fixed, accessible and rational that they can use to guide their work" (Bloxham 2011)



"assessment decisions at this level are so complex, intuitive and tacit that variability is inevitable" (Bloxham 2016)

Assessment Components vs Criteria

The grades of the CMS have qualitatively different requirements:

- it shouldn't be possible to achieve an "excellent" simply by doing more of the same kind of work required for a "good"

Summing components

- Simply summing the components of an assessment is unlikely to produce an appropriate result (at least not in any transparent way) ...
 - $Q1 (10 \text{ marks}) + Q2 (10 \text{ marks}) + \dots = \text{some mark out of } 100$
- Statistically manipulating the resulting profile is not meaningful

Criteria

- We can award marks for criteria instead of components:
 - has the student demonstrated the use of some proof technique not covered in the course?
 - is there an appropriate use of comments in the code ?
- These can be related more easily to the CMS grade requirements
- And they provide clearer feedback & connection to the learning outcomes

Combining criteria

Additive combination

- One approach to "criteria-based" marking simply weights and add the criteria
- But this does not take account of the qualitative difference required for the different grades
- And it is *compensatory*
 - to meet the learning outcomes, we probably don't want the student to pass if they have failed some "core" requirements, no matter how good the "extras" are

Decision rules

- Instead, we can give explicit decision rules
- These define what is required for each grade - for example ...
 - to obtain a "good" grade, all of the "code functionality" criteria must be good, and most of the "code readability" criteria must be good
- The failing rules also provide some explicit feedback

A motivating example

Basic criteria

- Understanding of the problem
- Completion of the project
- Quality of the work
- Quality of the report

Additional criteria

- Knowledge of the literature
- Critical evaluation of previous work
- Critical evaluation of own work
- Justification of the design decisions
- Solution of any conceptual problems
- Amount of work

UG4 & MSc Projects

should be derived from
the learning objectives

Example decision rules

0-19 Bad Fail: The project is inadequate on all of the basic criteria.

20-29 Clear Fail: The project is inadequate on more than one of the basic criteria, but not all.

30-39 Marginal Fail: The project is inadequate on one of the basic criteria.

40-49 III: The project is adequate on all of the basic criteria.

50-59 II.2: The project is at least fair on all of the basic criteria and is fair on most of the additional criteria.

60-69 II.1: The project is at least good on all of the basic criteria and is at least fair and sometimes good or excellent on all of the additional criteria.

70-79 Low I: The project is good or excellent on all of the basic and additional criteria; or it almost achieves this by being fair on only one of the additional criteria, and also has elements of the exceptional criteria.

80-89 High I: The project is good or excellent on all of the basic and additional criteria and also has elements of the exceptional criteria.

90-100 Outstanding I: The project is excellent on all of the basic and additional criteria, and has strong elements of the exceptional criteria.

PMark

Core functionality

- Accept a simple CSV file with criteria values for each student
- Evaluate a set of decision rules expressed explicitly in a simple language
 - support potentially large numbers of small rules
(this is hard to do automatically in any other way)
- Interpolate values between the rules to return a finer-grained result
- Support incremental development and adjustment of the rules to produce an "authentic" assessment

Extras

- Provide form-based input of the criteria values
- Generate textual "explanations" with reasons for the resulting grade
- Generate various plots and results from specified subsets of criteria
- Interactive exploration of the evaluation process
- GUI or command line

examples later ...



**Some things
we have learned**

Some things we have learned (1)

Teachers and students find this very "different"

- There is no "accumulation of credit" and no "weighting"
 - this requires a different "way of thinking" and careful explanation
 - approaching it with an "additive" mindset tends to be frustrating & unsuccessful
- Preparing a mark scheme forces the teacher to think about the criteria and how they relate to the learning outcomes
 - this takes longer
 - and it might even force a rethink of the assessment
 - but it pays-off later, especially for large and repeated assessments
- The rules are strictly non-compensatory - so it is easy for a relatively minor criterion to have a disproportionate effect on the result
 - accommodating this involves being explicit about the required "leniency"
- The best approach seems to be starting with the "high-level" requirements for a grade and refining them by identifying explicit criteria which demonstrate them

Some things we have learned (2)

Lots of questions and small ranges are good

- Questions with multiple dimensions should be split into "atomic" questions to avoid different markers balancing them differently
 - "is it clear & concise?" => "is it clear?" & "is it concise?"
- Markers have difficulty distinguishing between values on longer scales
 - is it "very good" or just "good" ?
- We now prefer a 4-point Likert scale for most questions:
 - "definitely no", "I don't think so", "I think so", and "definitely yes"
- Lots of small questions helps to average out uncertainties
- Even though there may be a lot of questions, this is easier to mark
 - we don't have to ponder whether this is "very good" or just "good"
- The failed questions can be used to generate very concrete suggestions on what would be necessary in order to achieve a higher grade

Some things we have learned (3)

Iterative & Flexible are good ...

- The mark scheme can be developed iteratively - starting with the main learning objectives and refining this into more detail
 - (PMark allows the effect to be explored using a set of dummy attribute values)
- We give the students the main criteria, but not the details
 - to avoid students "box ticking" without understanding
 - and so that we can ...
- Be prepared to change rules retrospectively
 - this allows us to cater for aspects of the assessment which clearly did not function as intended
 - or to acknowledge good solutions which use an unexpected approach

Estimates are helpful ...

- We found it useful to ask the marker to provide an intuitive value for the overall result
 - if this differs significantly from the computed result, it might indicate a problem with the rules.

Some things we have learned (4)

Being explicit is good ...

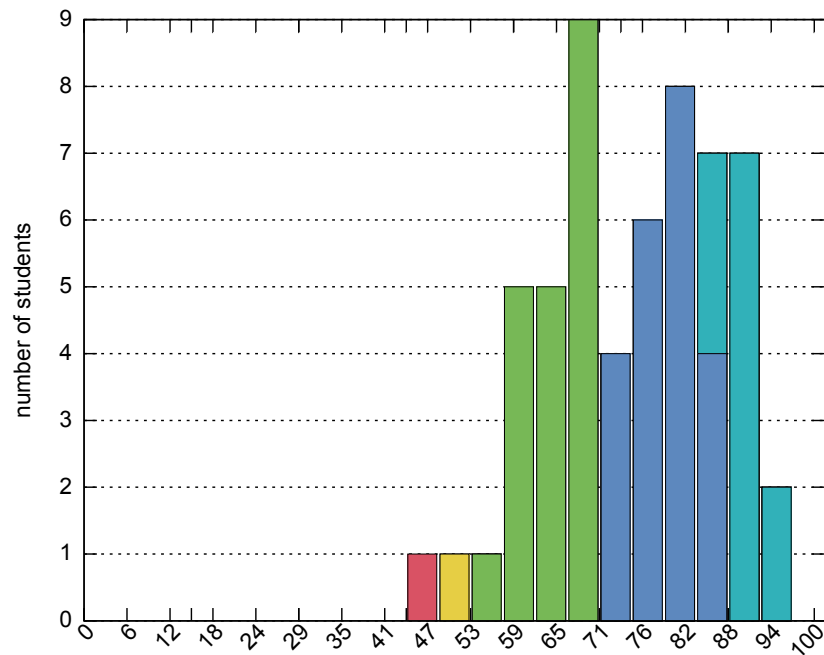
- Using more holistic criteria is sometimes useful - but it is important to be explicit about them. For example ...
 - "is there something exceptional about this submission? (explain in comments)"
- Similarly in terms of leniency ...
 - "a pass requires all of these criteria to be adequate and most of them to be good"
- Being explicit about these ensures that they are applied consistently and transparently

A bell-curve distribution is not necessarily good ...

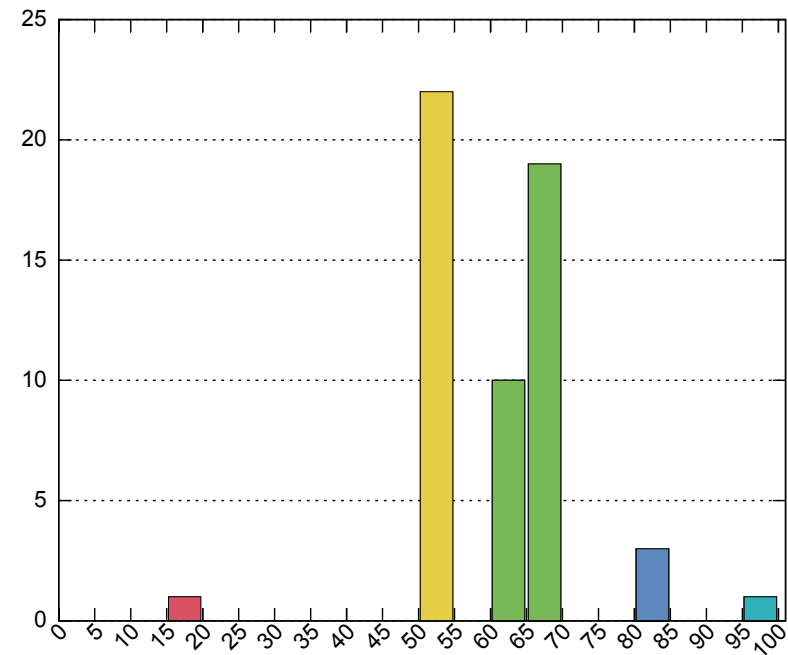
- A distribution with multiple peaks around the mid-grade points shows that the marker is making a clear distinction between the grades
 - this avoids clusters of results close to the grade boundaries

Mark distribution

Additive



Rule-based



Depending on how the criteria are combined and interpolated, the final mark distribution is likely be different from an additive scheme. For example:



PMark

- Probably not to be relied on for long-term support
- But I would be interested and happy to help anyone who wanted to try it out



◀ more ▶

 Scheme

ex8 ▾

 Data

form-test ▾

 Result

all ▾

 Graph

final-grade ▾

```
[attributes]
ex1: pass-fail #ex (
  question = "Did the student pass exercise 1"
  explain/pass = "complete exercise 1 satisfactorily"
)
ex2: pass-fail #ex (
  question = "Did the student pass exercise 2"
  explain/pass = "complete exercise 2 satisfactorily"
)
tutorials: out-of-10 (
  question = "How many tutorials did the student
attend"
  explain/10 = "attend all of the tutorials"
  explain = "attend more tutorials"
)

[rules]
pass: tutorials=5
pass: ex1=pass or ex2=pass

distinction: pass
distinction: ex1=pass
              and ex2=pass
```

 scheme 'ex8' computed using 'form-test' data



Upload



Download



New



Delete



Zip



Random



Forms



Compute



Save



PMark



◀ more ▶



ex8 ▾



form-test ▾



all ▾



final-grade ▾

id	ex1	ex2	tutorials
Adeen	pass	pass	10
Aliyah	pass	pass	8
Alyssa	pass	pass	7
Arianna	pass	pass	4
Ashton	pass	fail	8
Aurora	pass	pass	9
Bonnie	pass	pass	9
Dean	fail	pass	8
Eliza	fail	fail	7
Esther	pass	fail	4
Ethan	pass	pass	8
George	pass	fail	6
Jose	pass	pass	6
Justin	pass	pass	1



Upload



Download



New



Delete



Zip



Random



Forms



Compute



Save



scheme 'ex8' computed using 'form-test' data



My Form : Dean



Adam •

Elena •

Exercises

<input type="checkbox"/>	Did the student pass exercise 1 ?	fail ▾				
<input checked="" type="checkbox"/>	Did the student pass exercise 2 ?					
	<table border="1"> <tr> <td>grade</td> <td>to achieve this grade ...</td> </tr> <tr> <td>pass</td> <td>complete exercise 2 satisfactorily</td> </tr> </table>	grade	to achieve this grade ...	pass	complete exercise 2 satisfactorily	? ▾
grade	to achieve this grade ...					
pass	complete exercise 2 satisfactorily					

Tutorials

<input type="checkbox"/>	How many tutorials did the student attend ?	5 ▾
--------------------------	---	-----

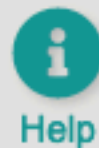
Feedback

B <i>I</i> <u>U</u>

new form for 'Dean'



PMark



◀ more ▶

Scheme

paulmc-pa ▾

Data

paulmcl2 ▾

Result

all ▾

Graph

DD ▾

id	DD	final-mark	pa-mark	my-report	pa-report	student-report	my-form
Alexandra	●	100	100				
Anna	●	97	97				
Arthur	●	86	86				
Ayden	●	78	78				
Caden	●	67	67				
Dakota	●	78	78				
Emilio	●	76	76				
Fiona	●	77	77				
Jonah	●	67	67				
Jose	●	75	75				
Leia	●	66	66				
Melanie	●	75	75				



Upload



Download



New



Delete



Zip



Random



Forms



Compute



Save



scheme 'ex8' computed using 'form-test' data



PMark



◀ more ▶

Scheme

SRCF-all ▾

Data

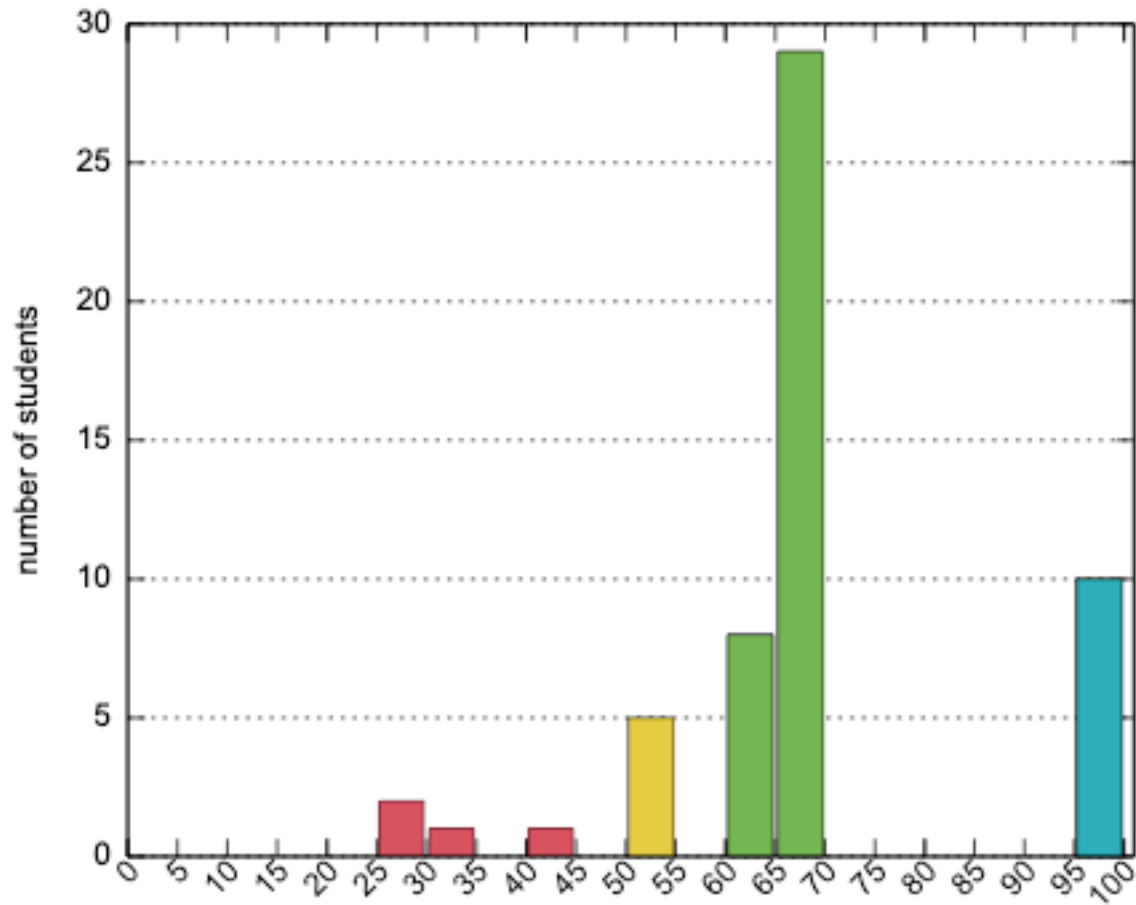
SRCF-mod ▾

Result

all ▾

Graph

poster-plot ▾



OK scheme 'ex8' computed using 'form-test' data



Upload



Download



New



Delete



Zip



Random



Forms



Compute



Save



PMark



◀ more ▶

Scheme

SRCF-all ▾

Data

SRCF-mod ▾

Result

all ▾

Graph

poster-plot ▾

- > rule report.marginal-pass
- > rule report.clear-pass
- > rule report.distinction
- > rule report.exceptional
- > subrule excellent-abstract
- ▾ subrule excellent-content ...
 - ▾ 2/2 criteria required ...
 - good-content and allbut 2 #content = yes
 - ↑ rule good-content
 - ▾ 11/13 criteria required ...
 - allbut 2 #content = yes
 - > body-analysis = yes-but
 - > body-background = yes-but
 - > body-conclusions = yes
 - > body-evidence = yes
 - > body-future = yes
 - > body-limitations = yes-but
 - > body-methodology = yes-but



- Upload
- Download
- New
- Delete
- Zip
- Random
- Forms
- Compute
- Save

OK scheme 'ex8' computed using 'form-test' data



◀ more ▶

Scheme

SRCF-all ▾

Data

srcf-rand ▾

Result

all ▾

Graph

total-plot ▾

Marks for Bryson report: 65%



Bryson achieved a clear-pass (55) for the REPORT.

Suggestions for a higher grade ...

For a distinction (70) ...

1. Discuss the significance of the findings more.
2. Use the Harvard referencing style.
3. List the authors' surnames in alphabetical order.
4. Use a more coherent writing style.
5. Explain the significance of the findings more clearly.
6. Explain the research methodology in more detail.
7. Integrate the sources more into the report.
8. Provide a full reference list at the end.
9. provide more recent sources (or explain why this is not possible).

Individual criteria

1. Are the main findings clearly summarized ? **yes**
2. Are the main conclusions clearly summarized ? **yes**
3. Is the background to the research presented ? **yes**



scheme 'ex8' computed using 'form-test' data



scheme 'SRCF-all' computed using 'srcf-rand' data



Upload



Download



New



Delete



Zip



Random



Forms



Compute



Save

**Questions?
Comments?
Discussion ... ?**

Paul Anderson
<dcspaul@ed.ac.uk>

PMark Homepage & Informatics PMark server

<https://pmark.nine3.org/server/public/home.html>

<https://ease.sweb.inf.ed.ac.uk/dcspaul/pmark/master/pmark.cgi/server>

Some References

- D. R. Sadler. Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(2):175–194, 2005.
- S. Bloxham, P. Boyd, and S. Orr. Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36(6):655–670, 2011.