



Resource: The CARFE principles of mark scheme design

In the table below is an overview of five principles for mark scheme design introduced in the 'Understanding and optimising your mark schemes' webinar.

These principles are connected to each other and aim to guide your decisions making for different types of mark scheme, with a particular focus on levels-based mark schemes.

In the 'actions to consider' column are some practical options that relate to each principle. These may not always be appropriate and other actions might be possible, but they aim to offer a way to begin thinking about connecting the principles to practical solutions.

Principle	Description	Actions to consider
C – Connect the item and the mark scheme	<p>The success of an item (from a validity perspective) depends upon making justified decisions about the assessment tasks to use.</p> <p>The content of the item and several other factors will influence how students respond. Student responses may be influenced by:</p> <ul style="list-style-type: none"> • Expectations of the student (i.e. similar questions done in the past) • Command words used (e.g. describe, evaluate etc.) • Paper layout (e.g. answer space) • Response prompts (e.g. whether the number of desired responses has been labelled) 	<ul style="list-style-type: none"> • Develop item and mark scheme drafts simultaneously • Think about how the command words influence how students interpret the question • Think about how many levels of response that you could realistically differentiate. Research suggests five is towards the top end of differentiation (Bramley, 2001)
A – Anticipate the range of student responses	<p>In many cases, items and mark schemes are designed before any testing with students. This can create circumstances where a student answers a question in an unanticipated way. In many cases unexpected responses can be catered for in the mark scheme during the standardisation process.</p>	<ul style="list-style-type: none"> • Think about what responses from students are ‘logically derivable’ from the item. • Have a second person (e.g. a subject specialist) answer the items as if they are a student. Did they answer in the anticipated way? • Are there any ways that you can pre-test the items without risking the security of the assessment or item?
R – Reduce marking complexity where possible	<p>The structure of the item and the mark scheme influence the cognitive processes underlying examiners’ marking behaviours (Suto & Greateorex, 2008; Suto & Nadas, 2008). Examination items that require examiners to scrutinise unexpected responses from candidates, or require examiners to evaluate a response using knowledge from several sources were more likely to result in lower reliability, compared to simpler items.</p>	<ul style="list-style-type: none"> • Provide <i>item-specific</i> guidance to support general levels criteria • Refer to key points in the mark scheme during a standardisation process • Reduce the number of marks within levels • Define what is not worth of credit

F – Format the mark scheme to facilitate marking quality	Mark schemes are the primary point of reference for examiners, and thus need to have a high degree of usability. Research has found that small changes to the formatting of mark schemes can improve usability ratings and overall marking quality (Child, Munro & Benton, 2015).	<ul style="list-style-type: none"> • Think about the use of a mark scheme as a working document – either printed or on screen. Can you make important parts of the mark scheme more visible? • Reduce the overall size of the mark scheme to one page for easy review of the levels in a level-based mark scheme. • Include indicative content alongside the level or levels of response it refers to. • Use bolding of key terms to maintain their salience.
E – Evaluate the performance of the mark scheme	<p>Investigating item-level data can reveal potential issues worthy of further investigation. For example, you may find that marks are ‘bunched’ in the middle of the mark range, or that some marks appear to be under-utilised. This does not necessarily mean that there is an issue with the mark scheme specifically, but will help you plan discussions with people using the mark scheme to identify any issues.</p> <p>It is also possible to use IRT-based statistical analysis to identify under-utilised marks (see Hughes & Shaw, 2016, for an example)</p>	<ul style="list-style-type: none"> • Consider the mark distribution data for your items. Does it appear that some marks are being over or under-utilised? • Check if the full range of marks within each level are being used. If not, this might suggest that examiners are not sure how to select a mark within levels. • Provide written guidance in the mark scheme about how to select a mark within a level.

References

Ahmed, A., & Pollitt, A. (2011). Improving marking quality through a taxonomy of mark schemes. *Assessment in Education: Principles, Policy & Practice*, 18(3), 259-278, DOI: 10.1080/0969594X.2010.546775.

Baird, J-A., Hayes, M., Johnson, R., Johnson, S. & Lamprianou, I. (2013). Marker effects and examination reliability: a comparative exploration from the perspectives of generalizability theory, Rasch modelling and multi-level modelling. <http://ofqual.gov.uk/files/2013-01-21-marker-effectsand-examination-reliability.pdf>

Black, B., Suto, I., & Bramley, T. (2011) The interrelations of features of questions, mark schemes and examinee responses and their impact upon marker agreement. *Assessment in Education: Principles, Policy & Practice*, 18(3), 295-318, DOI: 10.1080/0969594X.2011.555328

Bramley, T. (2001). The Question Tariff Problem in GCSE Mathematics. *Evaluation & Research in Education*, 15(2), 95-107.

Bramley, T. (2009). Mark scheme features associated with different levels of marker agreement. *Research Matters: A Cambridge Assessment Publication*, 8, 18-23.

Child, S. F. J., Munro, J., & Benton, T. (2015). An experimental investigation of the effects of mark scheme features on marking reliability. Retrieved from <https://www.cambridgeassessment.org.uk/Images/417277-an-experimental-investigation-of-the-effects-of-mark-scheme-features-on-marking-reliability.pdf>

Elliott, V. (2017). What does a good one look like? Marking A level English scripts in relation to others. *English in Education*, 51(1), 58-75.

Fowles, D. (2009). How reliable is marking in GCSE English? *English in Education*, 43(1), 49-67.

Hughes, S. & Shaw, S. (2016). Why do so few candidates score 4 out of 8 on this question? The issue of under-used marks in levels-based mark schemes. *Research Matters: A Cambridge Assessment publication*, 21, 42-48.

O'Donovan, N. (2005). There are no wrong answers: An investigation into the assessment of candidates' responses to essay-based examinations. *Oxford Review of Education*, 31(3), 395-422.

Suto, I., & Greatorex, J. (2008). What goes through an examiner's mind? Using verbal protocols to gain insights into the GCSE marking process. *British Educational Research Journal*, 34(2), 213-233.

Suto, I., & Nadas, R. (2008). What determines GCSE marking accuracy? An exploration of expertise among maths and physics markers. *Research Papers in Education*, 23(4), 477-497.

Tisi, J., Whitehouse, G., Maughan, S., & Burdett, N. (2013). A review of the literature on marking reliability research. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/605688/0613_JoTisi_et_al-nfer-a-review-of-literature-on-marking-reliability.pdf

Williamson, J., & Child, S. F. J. (2019). *How can mark scheme design support reliable and valid school-based assessment?* Presented at the European Conference of Educational Research, Hamburg, Germany, 2-6 September 2019.