

26 September 2024

[REDACTED]

Tēnā koe [REDACTED]

Official Information Act Request

Thank you for your request of 7 September 2024, under the Official Information Act 1982, for the following information:

Therefore, I request, under the Official Information Act 1982, all documents and correspondence related to the following queries:

1. *How is a candidate's response assessed against the marking schedule?*

a. Additionally, is this done by a human, a computer, or a combination? (I note this might differ across each standard)

2. *Following this, how is a grade awarded based on the criteria achieved?*

a. As part of this, what number or percentage of criteria are required for an Achieved grade?

b. Can a candidate's grade differ based on other responses? In other words, is a bell curve (or similar tool) used as part of this process?

c. Additionally, is this done by a human, a computer, or a combination?

I request this information in respect of US32405 (writing), US32403 (reading), and US32406 (numeracy), and in respect of both 2023 assessment events and both 2024 assessment events. I understand that you are not necessarily obligated to directly answer my questions, but I have set my request out in this way to make it quite clear what I am requesting.

Your request has been considered under the Official Information Act 1982.

How is a candidate's response assessed against the marking schedule for US32405, US32403, and US32406 in respect of both 2023 assessment events and both 2024 assessment events? Is this done by a human, a computer, or a combination?

32403: auto marked by machine. Each question is either correct or incorrect.

32405: Tasks 1 and 2 are marked by humans. In the bench-marking process, a small group of senior markers select and annotate a set of exemplars for both tasks using the assessment schedule. These exemplars are used in the marker panel meeting, where all markers are trained to use the schedule and the exemplars to mark through a 'single pair of eyes'. Every marker is assigned a check marker and there is regular check-marking of all markers throughout the marking

period. Each piece of writing is assessed on four aspects – content, structure, language choices and technical accuracy. Evidence for each aspect is assessed as one of the following: none; minimal; weak; sufficient; strong.

32405: Task 3 is auto marked by machine. Each question is either correct or incorrect.

32406: marked by humans. In the bench marking process, a small group of senior markers finalise the assessment schedule, then select and annotate a set of exemplars. These exemplars are used in the marker panel meeting. In this meeting all markers are trained to use the schedule and the exemplars to ensure marking is done through a 'single pair of eyes'. Each marker is assigned a check marker and there is regular check marking of all markers throughout the marking period. Each question is either correct or incorrect.

Following this, how is a grade awarded based on the criteria achieved?

a. As part of this, what number or percentage of criteria are required for an Achieved grade?

All standards follow a rigorous four phase process for determining Achieved or Not Achieved results. This includes:

- Phase one: Consideration of the assessment instrument
- Phase two: Consideration of marking candidate responses
- Phase three: Consideration of candidate performance
- Phase four: Evaluation of the assessment to decide upon a cut score

The four-phase process provides a comprehensive data analysis of the assessment items, which is used to determine a cut score for each assessment.

For 32405 and 32406, the decision for the Achieved / Not Achieved cut score is made by humans using data analysis tools including the Angoff method. For 32403, the Bookmarking method is used.

The cut score may vary from assessment to assessment as the data from each assessment needs to be analysed before that decision can be made.

There must be at least some evidence for all outcomes of the standard and sufficient evidence across all outcomes to meet the requirements of the standard.

See Appendix 1 (attached) for the process for setting a Literacy and Numeracy cut score.

b. Can a candidate's grade differ based on other responses? In other words, is a bell curve (or similar tool) used as part of this process?

No. As these are unit standards, if students provide sufficient evidence to meet the requirements of the standard they will achieve.

c. Additionally, is this done by a human, a computer, or a combination?

For 32403, 32405 and 32406 the decision for the Achieved / Not Achieved cut score is made by humans using the data analysis tools referred to above.

Your response may be published on our website after five working days. Your name and contact details will be removed before publication.

If you require further assistance or believe we have misinterpreted your request, please contact Elizabeth Templeton in the Office of the Chief Executive, email elizabeth.templeton@nzqa.govt.nz or telephone (04) 463 3339.

You have the right to seek an investigation or review by the Ombudsman of this decision under section 28(3) of the Official Information Act 1982. Details of how to make a complaint can be found

at www.ombudsman.parliament.nz. You can also telephone 0800 802 502 or write to the Ombudsman at PO Box 10152, Wellington, 6143.

Nāku nā

A handwritten signature in black ink, appearing to read 'Grant Klinkum', written in a cursive style.

Dr Grant Klinkum
Pouwhakahaere/Chief Executive

Appendix 1: Process for setting Literacy and Numeracy cut-score

Phase one: Consideration of the assessment instrument

Phase two: Consideration of marking candidate responses

Phase three: Consideration of candidate performance

Phase four: Evaluation of the assessment to decide upon a cut score

Phase One: Consideration of the assessment instrument

Development Team	<i>Highlight any issues from development that could have impacted upon scores</i>
Experience and background of development team	
Changes in development team	
Issues with development due to authoring tool	
Changes/continuities in the instrument from previous years	
Rationale for any changes	
Number of objective and subjective items	
Format of the assessment	
Identification of the construct through a matrix that relates the curriculum and achievement objectives (written form) at the relevant level to question/items. Including intended assessment outcome for each item	
Number of question items	
Scoring format per item	
Weighting attached to items	
Rationale for weights	
Estimate of the difficulty of assessment relative to difficulty in previous years	

Phase Two: Consideration of marking candidate responses (Writing 32405 and Numeracy 32406 only as reading auto marked)

Marking Team	<i>Highlight any issues from marking that could have impacted upon scores</i>
Composition of team	
Previous experience of team	
Changes to team from previous years	
Changes to marking process	
Impact of the number of entries and responses on marking	
Marking duration	
Marker reliability and validity process	
Frequency of marker quality control sampling	
Ratio of accurate to inaccurate sampled responses (for panel and per marker)	
Monitoring process and data available during marking	
SOP for each marker for each question item desirable but may not be possible	
Issues noted regarding the fit of the candidate responses to expected response	
Changes to schedule and weighting during marking	
Estimate of the difficulty of the assessment relative to previous years	

Phase three: Consideration of candidate performance

NAF in consultation with SME	
Issues from authoring, sitting, or marking application that could have impacted results	
Entries, voids, and absences for assessment	
Question/item data, percentage of candidate subs that responded, percentage at score	
Candidate year level and e-asTTle data	
Angoff analysis data	

State the claim NZQA makes for the assessment, and highlight any issues from instrument development, marking, or candidate performance that the National Assessment Facilitator considers could have impacted upon scores.

Provide a recommended cut-score:

Attach any appendices.

Phase four: Evaluation of the assessment to decide on cut-score

Cut-score Team	
Changes in cohort	
Changes in characteristics of the submission	
Effects of instrument on scores	
Relationship of assessment instrument to the construct and the claim	
Distributions and variances of results in the current year and previous years	
Impact of the assessment and possible cut-scores on candidate outcomes	

Final cut-score confirmed	
Final cut scored changed (justified)	