Research Matters / 36
A Cambridge University Press & Assessment publication

ISSN: 1755-6031

Journal homepage: https://www.cambridgeassessment.org.uk/our-research/all-published-resources/research-matters/

An example of redeveloping checklists to support assessors who check draft exam papers for errors

Sylvia Vitello, Victoria Crisp and Jo Ireland (Research Division)

To cite this article: Vitello, S., Crisp, V., & Ireland, J. (2023). An example of redeveloping checklists to support assessors who check draft exam papers for errors. *Research Matters: A Cambridge University Press & Assessment publication, 36*, 46–58. https://doi.org/10.17863/CAM.101744

To link this article: https://www.cambridgeassessment.org.uk/Images/research-matters-36-an-example-of-redeveloping-checklists-to-support-assessors-who-check-draft-exam-papers-for-errors.pdf

Abstract:
Assessment materials must be checked for errors before they are presented to candidates. Any errors have the potential to reduce validity. For example, in the most extreme cases, an error may turn an otherwise well-designed exam question into one that is impossible to answer. In Cambridge University Press & Assessment, assessment materials are checked by multiple assessment specialists across different stages during assessment development. While human checkers are critical to this process, we must acknowledge that there is ample research showing the shortcomings of being human (e.g., we have cognitive biases, and memory and attentional limitations). It is important to provide assessment checkers with tools that help overcome or mitigate these limitations.

This article is about one type of checking tool – checklists. We describe a research-informed, collaborative project to support assessors in performing their checks of exam papers. This project focused on redesigning the instructional, training and task materials provided to assessors. A key part of this was to design checklists for assessors to use when performing their checks. In this article, we focus primarily on the approach that we took for these checklists in order to draw readers' attention to the complexity that is involved in designing them and to provide a practical example of how research can be used strategically to inform key design decisions.

Cambridge University Press & Assessment is committed to making its documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team:
Research Division, researchprogrammes@cambridgeassessment.org.uk
If you need this document in a different format contact us, telling us your name, email address and requirements and we will respond within 15 working days.

© Cambridge University Press & Assessment 2023
Full Terms & Conditions of access and use can be found at
T&C: Terms and Conditions | Cambridge University Press & Assessment

An example of redeveloping checklists to support assessors who check draft exam papers for errors

Sylvia Vitello, Victoria Crisp and Jo Ireland (Research Division)

Introduction

When new exam papers are drafted, they need to go through a quality assurance process, just as we would expect for all important educational and non-educational products. Many aspects of question design contribute to ensuring that exam results accurately reflect a learner's relevant knowledge, skills and understanding and, thus, to ensuring that it is appropriate to use the assessment results in the intended ways. For example, features such as language accessibility, visual resources and context affect how well learners can show what they have learned (e.g., Crisp & Sweiry, 2006; Ahmed & Pollitt, 2007; Crisp, 2011; Crisp & Macinska, 2020). The most extreme problems with exam questions occur where a clear error appears in a paper. For example, this could be a factual inaccuracy which then renders a question unanswerable, or something on an exam paper that gives away the answer to a question elsewhere on the paper. It is of great importance that awarding bodies have robust procedures in place to ensure that questions are of high quality and errors are avoided. These procedures often involve a staged process through which exam papers are developed incorporating input from a number of assessors with expertise in the relevant subject.

Recently, Suto and Ireland (2021) reviewed the literature from the field of error detection and explored the psychological and system-level causes of errors in order to recommend a set of principles for how to minimise errors in exam papers. They highlighted various psychological causes of errors that are relevant to the context of exam paper construction. These include cognitive biases and limitations related to memory, attention and our tendency to use heuristics (i.e., imperfect, non-rational methods) over analytical approaches during tasks involving judgement and decision making. These human characteristics can cause us to make errors and to fail to detect ones made by others or ourselves. A checklist is one type of tool that can potentially help avoid errors by supporting appropriate checks during the stages of a process. Gawande (2010) explains how the use of checklists in error prevention or detection is supported by psychological theories of cognition and attention. Checklists can help overcome or mitigate psychological

© Cambridge University Press & Assessment 2023

error factors by acting as aide-memoires that encourage the user to take a more systematic and analytic approach to the checking task. Another reason why checklists can help with certain cognitive limitations is that they specify the checker's task, making it clear to the checker what they should check for. Under-specification of a task or process has been argued to be a significant factor in the production of human error across industries (Reason, 2013).

This article is about some of the checklists that are used to support exam paper production processes at Cambridge University Press & Assessment. We discuss the approach that was taken to redevelop these checklists. The aim of this article is to draw attention to the complexity that is involved in designing checklists and, also, to provide a practical example of how research can be used strategically to inform key design decisions.

Project context and aim

In Cambridge University Press & Assessment, exam papers and other assessment materials are produced through a process of drafting, review and refinement involving a number of professionals with appropriate subject and assessment expertise. Focusing more specifically on OCR (one of our exam boards), during its assessment materials production process, exam papers (and other assessment materials) undergo a specific, carefully designed series of checks after a complete draft has been produced. These checks are aimed at ensuring the quality of the assessment materials, including detecting errors so that they can be corrected before the paper is sat by candidates. This article describes a research-informed project that involved redeveloping the checklists (and other related materials) to support OCR's assessors at this stage of the quality assurance process when a complete draft of the paper has been produced. The focus was on four of OCR's checking roles:

- Candidate Proxy – an assessor who works the exam paper as if they are a candidate.
- Assessment Marker – an assessor who marks the Candidate Proxy's exam script and reviews the alignment between the exam paper and the mark scheme.
- Assessment Analyst – an assessor who applies a question analysis technique to all of the questions in the paper, whereby they deconstruct the constituent words, phrases and parts of the question.
- Pre-exam Check – one of the final checks of exam papers, which is performed by an assessor[1] whose primary aim is to catch any serious errors that could affect the candidates' ability to answer questions.

The Candidate Proxy, Assessment Marker and Assessment Analyst all complete their checks around the same time. The results of these checks are then reviewed by the assessment manager within OCR and the paper is revised as needed. Some papers then undergo a proofreading check and a plagiarism check followed by another internal review by the assessment manager. The Pre-exam Check occurs after all these other stages.

1 The Pre-exam Check is sometimes performed by an external assessor or an internal assessment specialist with no prior involvement in the exam paper's development.

© Cambridge University Press & Assessment 2023

Before the redevelopment project started, OCR had already been providing their assessors with a type of checklist to use when performing these checking roles. These checklists took the form of a set of questions about the exam paper for assessors to consider and respond to when performing their checks. These checklist questions were embedded within a document known as the "report form". Assessors used this report form alongside conducting their checks and filled it in with details of their evaluation of the paper including the issues they identified. These completed forms were then reviewed by the assessment manager.

The main focus of the redevelopment project was to create a new report form for each of these roles, which would contain a new checklist that was strategically designed to ensure that all issues appropriate to a role would be checked. In addition to the checklists and report forms, the other materials that checkers would be provided with were reviewed and revised (specifically, instruction documents and training materials) to ensure that they also cohered with the new report forms. This article reports primarily on the checklists and report forms. The final checklist for the Assessment Analyst role is shown in the Appendix as an example.

Overview of the redevelopment approach

As this project was strongly linked to the operational running of the assessment materials production process and aimed to take a research-informed approach, a cross-department working group was established including researchers, assessment managers, a manager who oversaw a team of staff who co-ordinate the work of external assessors, and a project manager. This collaborative approach was critical for ensuring that, while drawing on the relevant research literature, decisions about the new checklists and redeveloped materials aligned with OCR's vision and intentions for the checking roles, and that the materials would be useable in practice.

The project was collaborative and iterative with the redevelopment usefully informed by input gathered at various stages of the process from a range of people with different roles in the question paper process. The main stages of work were:

- a mapping exercise in which a taxonomy of question paper error types was mapped against each of the four checks in order to set out what should be checked for each role
- initial design and drafting of checklists and report forms
- consulting internal staff involved in the assessment materials production process, followed by refining the materials as needed
- piloting the materials in several subjects with those who conduct the checks – assessors were asked to check an example exam paper using the revised checklist and other support materials and provided feedback, after which the materials were further refined
- consulting internal staff again on the changes and minor further revisions.

© Cambridge University Press & Assessment 2023

The next section of this article describes the mapping exercise that formed the foundation of our checklists. The two subsequent sections focus on design decisions relating to the checklist items and design decisions about the report forms within which the checklists would appear. These design decisions evolved throughout the course of the redevelopment project, with final decisions often resulting from an iterative consideration of information, discussions and evidence across different stages of the project (e.g., initial design, consultation, piloting and refinement process). For brevity, the sections on design decisions bring together themes that arose at any stage of the redevelopment process rather than separating out points based on the chronology of events.

Redeveloping the checklists and report forms

Mapping exercise

As the starting point for revising the checklists and report forms, OCR members of the project working group mapped out the types of error that each checking stage (role) was intended to identify. A taxonomy of 42 error types that had recently been developed by Suto et al. (2023) was used. This was derived from an analysis of Cambridge University Press & Assessment's records of assessment materials errors from across several years (2012 to 2018).

The mapping exercise was based on an approach recently developed by Suto et al. (2023) as a way to systematically and strategically evaluate existing checks of assessment materials. They showed how the approach can help assessment teams to understand, for example, whether all error types are being targeted across checking roles, how many error types are targeted by any individual role (i.e., checker workload), and how many times each error type is checked for (i.e., by one or multiple roles). Together, this can aid in building a strategic map of which error types *should* be checked by each role.

In our redevelopment project, the OCR members of the working group conducted a review of each check using the Task Descriptors,[2] recruitment criteria and the instruction documents given to assessors about the checking tasks, looking at which of the 42 error types were being targeted by each role according to these documents. This led into mapping out the intentions for the checking roles in terms of whether checkers in each role should be expected to look for and identify each type of error. The relevant working group members identified four categories to help them distinguish between different kinds of role intentions:

- Core focus – This error type is a core focus of this checking role. This means that the checker performing this role should conduct a thorough check for all errors of this type.
- Peripheral focus (high impact) – This error type is not a core part of this checking role and should not be a main focus for checkers. However, because of the nature of the role any *high* impact errors of this type should be

2 The Task Descriptor for each assessor role is a publicly available document containing a brief description of what is involved in the role. For example, the Task Descriptor for the Candidate Proxy can be found here: https://www.ocr.org.uk/Images/471234-candidate-proxy-task-descriptor.pdf.

© Cambridge University Press & Assessment 2023

identified if the checker performs their check correctly.

- Peripheral focus (lower impact) – This error type is not a core part of this checking role and should not be a main focus for checkers. However, because of the nature of the role it is possible (or even likely) that lower impact errors of this type would be identified if the checker performs their check correctly.
- Not expected – This error type is not intended to be detected by the checker performing this checking role.

These role intentions were recorded in a mapping grid of the kind shown in Table 1.

Table 1: An illustration of the kind of output produced from the mapping exercise (note: this example illustrates the concept and does not reflect the actual output for this project).

	Error type			
	1 – topic not on relevant syllabus	**2 – topic inappropriate for exam/ component**	**3 – item has a factual inaccuracy**	**4 – item is of an inappropriate level of demand**
Checking role 1				
Intention of role	Core focus	Core focus	Peripheral (high impact)	Peripheral (high impact)
On the checklist	Yes – explicit	Yes – explicit	Yes – explicit	Yes – implicit
Checking role 2				
Intention of role	Not expected	Peripheral (lower impact)	Core focus	Core focus
On the checklist	No	No	Yes – explicit	Yes – explicit

In addition, as can be seen in the example, the mapping grid was also used to record initial decisions about whether the error type should be addressed on the checklists being developed. For example, did it need an explicit checklist item, or could it potentially be an implicit part of a checklist item? These considerations were important because it would not be practical to ask all assessors to check for all error types in the Suto et al. (2023) taxonomy, given the high number of error types.

The mapping grid was regularly referred to during the design stage of the checklist development, as it helped the working group to make strategic and systematic decisions about what should be included in the checklists.

Design decisions about the checklist items

Many decisions had to be made about how to design the checklist items. This section draws attention to several different aspects of the checklists, focusing on ones where decisions were complex or had a particularly strong influence on other design decisions. The aspects discussed in this section relate to the type, length, content, phrasing and structure of the checklists.

© Cambridge University Press & Assessment 2023

The type of checklist

One fundamental decision that we needed to make was what type of checklist to have. In the literature, two main types of checklist stand out: Read-Do and Do-Confirm (Gawande, 2010), where the distinction concerns when the actions on the checklists are taken. Read-Do checklists are completed as part of the checking task, with the checklist items acting as prompts for actions and checkers marking the checklist items as complete as they go along. Do-Confirm checklists are completed after the tasks are done; checkers perform their tasks from memory and experience first and then use the checklist to confirm that all of the checklist items have been completed. Neither of these types of checklist seemed to fully reflect what checkers do in the exam paper context. Unlike in aviation or healthcare contexts, assessment checkers are not required to take direct action (i.e., they do not make any changes to the exam paper); instead, their task is to review and report on the state of the exam questions and provide recommendations. Therefore, it was important that the design of the checklists supported this different type of checklist.

A relatively simple way of achieving this was to reflect this task of reviewing and reporting in both the wording of the checklist items and in how checkers were asked to respond to the item. In line with Suto et al. (2023), we decided that checklist items for our exam paper checks should be phrased as questions. This communicates clearly what the checkers need to review the exam paper for and that they need to provide an answer to this question on the checklist. An example of the form of the checklist item is shown below:

> "Are all the answer spaces appropriate in terms of both type (e.g., lines, graph paper) and size?"

As that example shows, we also decided to phrase each checklist item as a "yes/no" question, specifically where "yes" meant that there were no problems of this kind with the exam materials. The aim of structuring all checklist items in this way was to make it easy and quick for those using the checkers' completed checklists (e.g., assessment managers) to see if any problems had been identified. Another option of "not sure or unable to say" was also provided to encourage checkers to record issues that they felt might be problems even if they were unsure.

Another debate during development was whether or not to ask checkers to record that they had checked every checklist item for each individual exam question or question part (e.g., by completing each cell of a grid showing each exam question part). This had to be considered carefully early in the redevelopment project, as it had implications for various fundamental aspects of the checklist (e.g., checklist length, structure, phrasing). Potential advantages of this "question-by-question" method were considered, which included that it could help draw the checkers' attention to each checklist item for each exam question part, in line with the "point and call" method[3] (Hikida et al., 2015), and that it

3 "Point and call" checks are used in various industrial contexts in Japan and involve use of a checklist to prompt pointing at the item to be checked and calling out its state. The method has been found to reduce error rates (Haga, Akatsuka & Shiroto, 1996, as cited in Hikida et al., 2015).

© Cambridge University Press & Assessment 2023

could provide more detailed data on where errors or issues occurred. Potential disadvantages were also discussed, including concerns that assessors may perceive this as added administration, and that it could result in a large matrix which might discourage assessors from actively engaging with it, especially for papers with a large number of questions and question parts (e.g., mathematics). The latter could lead to some assessors simply ticking boxes by default without actually checking each aspect carefully, negating the purpose of such a grid. The possibility of using a "question-by-question" strategy was also complicated by the fact that some checklist items were elements to be checked at the level of the whole paper rather than at the level of the question or question part. Due to these complications, it was decided not to require a record of the checking of each question or question part for each checklist item. However, it was made clear in all question-level checklist items and in accompanying instructions that all questions should be checked.

The length of each checklist

Another general factor that we considered early in the design stage was how long the checklist should be (i.e., how many checklist items). It was important to start discussing checklist length early in the process because it would affect other key design decisions such as the amount of content (i.e., how many errors and issues) that could be covered in the checklists and how to express this content in checklist items (e.g., should we have many specific items or fewer broader items?). Ultimately the lengths of the final checklists were the result of carefully considering and balancing different factors.

In the literature on safety industries' use of checklists, Gawande (2010) promotes short checklists, explaining that "a rule of thumb some use is to keep it between five and nine items, which is the limit of working memory" (p. 123) and to focus on the "killer items" – the most critical checks. It is important that checklist designers tailor guidelines to individual situations, assessing the impacts (positive and negative) that deviations from the guidelines may have on the checkers and their capability to complete their checks and any concurrent tasks they perform.

For our exam paper context, it was decided that our checklists could benefit from being longer and more comprehensive for a number of reasons relating to the context and purpose of these checklists, which differ from those in safety-critical industries. In particular, it was considered important to use the checklists as a means to ensure clarity around the remit of the checking roles, as under-specification of checking roles is a key factor in increasing the risk of error (Reason, 2013). Being flexible with regard to checklist length was deemed to be reasonable because of two other features of exam paper checks. The exam paper checks were not as strictly constrained by time limits as some of the checks for which the safety-industry advice was designed. In addition, the question paper checks were to be presented in a written document rather than having to be recalled from memory, which meant that the number of items on the checklist did not need to be constrained by memory limitations. Nevertheless, in designing the checklists, much attention was still paid to what was a reasonable number of points to expect the assessor to check for, bearing in mind the nature of their task. For example, the

© Cambridge University Press & Assessment 2023

Assessment Analyst's task involves careful analysis of the text and how different parts of this relate to one another. Therefore, it was considered undesirable to distract them from their focus on this task with a large number of checklist items. The final numbers of checklist items for each role ranged from 7 to 16.

Content covered across each checklist

Many design decisions were affected by views on what content the checklist should target (i.e., which paper errors). We considered many factors in our decision-making about the checklist content, including: the remit and scope of the checking role; the importance of the error for the checking role and assessment manager; and the purpose and usefulness of including a particular content point in the checklist.

A lot of this key information had been set out as part of the mapping exercise, and, therefore, the mapping grid was our starting point. As described earlier, the mapping grid distinguished between different error types in terms of their importance and function in the checking role. This facilitated the prioritisation of content, which was important given the discussions around checklist length. It was decided that, as a minimum, the checklist should include each error type that had been identified as a "core focus" of the checking role.

The mapping exercise only included the error types that had been identified by Suto et al. (2023). Therefore, it was also important to check with OCR colleagues (such as the assessment managers) whether it would be useful to include other checks in the checklist. Based on these discussions, a small number of checks were requested that did not originate from Suto et al.'s (2023) list of error types. For example, for the Candidate Proxy, Assessment Analyst and Assessment Marker roles, the assessment managers felt it was important for the first checklist item to relate directly to the key nature and purpose of the checker's task. Accordingly, for the Candidate Proxy and Assessment Analyst, the first checklist item focused on question answerability:

> "Are all the questions answerable (i.e., candidates who have studied the full course should be able to make a sensible attempt at answering each question)?"

For the Assessment Marker, the first item asked about alignment between the Candidate Proxy's answers and the question and mark scheme:

> "For each question, does the Candidate Proxy's response align with the question and the mark scheme?"

Content covered by individual checklist items

The decisions about content discussed in the previous section were about what to cover across each checklist as a whole. Decisions also needed to be made about how much content individual checklist items should cover. In practice, discussions about these decisions were often intertwined but we separate them here for readability.

© Cambridge University Press & Assessment 2023

One of the most important considerations with regard to the content of individual checklist items was how focused we needed each checklist item to be. Should items focus on one error type only? Again, this decision has many consequences. It would inevitably affect checklist length. If each checklist item were focused on one error type, then we would need more checklist items than if checklist items were broader in scope. However, the broader the checklist item, the less helpful it may be for the checkers, and for the assessment managers using the outputs of the checks.

In the final checklists, some checklist items were a direct representation of one error type (based on the mapping grid). This level of specificity was often chosen in cases where it was considered particularly important for checkers to focus on an error type or for assessment managers to have confirmation that the exam paper had been checked for a specific issue. For example, for error type 3 in Suto et al.'s (2023) taxonomy, "Item content factually inaccurate", a checklist item was written asking:

> "Is the subject content of all the questions factually accurate?"

In other cases, where error types were closely related, more than one error type was addressed by one checklist item. For example, error types 21 and 22 in Suto et al.'s (2023) taxonomy relate to inconsistency across the different parts of a question and inconsistency between different questions, respectively. For checking roles that were expected to check for both of those error types, one checklist item was written combining both aspects:

> "Does the paper avoid inconsistencies within and between questions (e.g., terminology, subject content)?"

Phrasing of the checklist items

The phrasing of checklist items was carefully considered at multiple times during the design stage given the importance of clearly communicating the check to the checker. As described earlier, one of these decisions was to phrase the checklist items as questions because it made the checking task and output of the task (i.e., yes, no, not sure or unable to say, not applicable) clear to the checker and to those who would need to review the completed checklists.

Another consideration was making sure that the phrasing clearly described the aspect of the exam question or paper that the checkers needed to review and make a judgement about. Guidance about preparing checklists argues that wording should be kept simple and exact, using words familiar within the context (Gawande, 2010). The project working group tried to follow this principle in the design of the checklists, and as part of consultation efforts, a staff member with relevant training conducted a review of the materials to ensure simplicity and conciseness in the language used. Phrasing decisions were not always straightforward, as there was sometimes a trade-off between simplicity and specificity. Although checklist guidelines recommend conciseness (Gawande, 2010), the guidelines and wider psychology literature on human factors in error

© Cambridge University Press & Assessment 2023

also emphasise the risk of ambiguity and under-specification (Reason, 2013). One strategy that we used was to include additional information within parentheses for checklist items where there was potential ambiguity, for example:

> "Are all the questions answerable (i.e., candidates who have studied the full course should be able to make a sensible attempt at answering each question)?"
>
> "Are all the answer spaces appropriate in terms of both type (e.g., lines, graph paper) and size?"

For each checklist, the content and wording for each of the checklist items for that role was drafted and refined through discussion with members of the project working group, and later through wider consultation and feedback from piloting.

Structure and layout of the checklists

Given the relatively high number of checklist items for most roles (7 to 16), consideration was given to whether the checklist items within a checklist could be grouped into sections, as proposed by Degani and Wiener (1993). Some checklist items were factors to be checked in each individual exam question (e.g., whether each question is clear, unambiguous and will not cause confusion), while others were elements to be checked at the paper level (e.g., whether the exam paper avoids testing exactly the same content in more than one place within the paper). For the Assessment Marker's checklist, there were also some marking-related checks (e.g., whether the mark scheme rewards candidates for what the questions ask for). These three areas (question-level, paper-level and marking-related) provided a logical way of dividing the checklists into sections to make them more manageable for assessors to use.

Design decisions about the report forms

In the checking processes already in place before the redevelopment project, checkers completed a report form alongside conducting their checks. This contained: a table for assessor and question paper details; basic instructions for using the form; issues to check for; and a comments table for recording details of issues found. For the redevelopment, new report forms were drafted drawing on the structure and information in the pre-existing report forms. The table for assessor and question paper details and the basic instructions were revised as needed, with a general aim of ensuring consistency between the basic elements of the report forms for different roles, unless there was a good reason for there to be differences.

The checklist items replaced the points to check for listed in the previous forms. The checklist items were presented in tables with a column containing the checklist items and a separate column within which checkers should respond Y for "yes", N for "no", ? for "not sure or unable to say", or N/A for "not applicable". Where relevant, the checklist items were separated into different tables for the sections on question-level, paper-level and marking-related checks.

The comments tables were also updated, ensuring consistency between roles,

© Cambridge University Press & Assessment 2023

where possible. A key design focus was to ensure that completed forms would provide useful information to assessment managers. In the pilot versions of the report forms for the Candidate Proxy, Assessment Marker and Assessment Analyst, the comments table included a column for recommendations to ensure that suggestions for improvement were provided alongside comments on potential issues. This was not included in the pilot report form for the Pre-exam Check, as it was initially felt to be unnecessary given the nature of the errors most likely to be identified at this late stage, which should automatically suggest the required correction. However, the experience of the pilot suggested that adding the recommendations column to the table would be useful to ensure clear information is provided by assessors to the assessment managers. Another decision to aid useability was to include a column in the comments tables for all checking roles where checkers are asked to record the checklist item to which each comment relates. The intention of this was to allow assessment managers to easily confirm that checkers had written in the comments table for each checklist item where they identified a possible issue. It was also considered to be a way to support data analysis of the kinds of issues identified during checks.

Summary and reflections

The project described in this article provides a case study of using research and guidance about checklist design to support a systematic redevelopment of the materials to support exam paper checking processes, intended to ensure well-designed assessments. For each checking role, checklists were carefully designed and integrated into forms for assessors to use when carrying out checks. Associated instructions and training materials were updated to cohere with the checklists and report forms. The establishment of a cross-department working group and wider consultation and piloting were valuable in drawing on broad expertise and ensuring usability of materials for those involved.

The materials developed in this project are now in use by assessors and assessment managers. Future review of their usefulness is planned so that they can be refined, if needed, to optimise checking processes. It is hoped that further work can explore how the data from the report forms could be used to support the analysis of errors that occur in live exam papers and, indeed, of errors that could have appeared in live exam papers but were successfully identified during checks. Feeding back to assessors on errors and "near misses" also has potential benefits. Such ongoing monitoring and feedback can help to continually improve processes and assessor expertise, thus ensuring exam paper quality, something of crucial importance to the accurate and fair assessment of learners.

Acknowledgements

We would like to thank the many OCR colleagues who contributed to the project, in particular Frances Wilson, Stephanie Wyre, Naomi Rowe, Kate Elliott, Helen O'Leary, Jane Bowen, and all the assessors who gave us feedback on the checking materials.

© Cambridge University Press & Assessment 2023

References

Ahmed, A., & Pollitt, A. (2007). Improving the quality of contextualized questions: an experimental investigation of focus. *Assessment in Education: Principles, Policy and Practice, 14*(2), 201–232.

Clay-Williams, R., & Colligan, L. (2015). Back to basics: checklists in aviation and healthcare. *BMJ Quality & Safety, 24*(7), 428–431.

Crisp, V. (2011). Exploring features that affect the difficulty and functioning of science exam questions for those with reading difficulties. *Irish Educational Studies, 30*(3), 323–343.

Crisp, V., & Macinska, S. (2020). Accessibility in GCSE Science exams – students' perspectives. *Research Matters: A Cambridge Assessment publication, 29*, 2–10.

Crisp, V., & Sweiry, E. (2006). Can a picture ruin a thousand words? The effects of visual resources in exam questions. *Educational Research, 48*(2), 139–154.

Degani, A., & Wiener, E. L. (1993). Cockpit checklists: concepts, design, and use. *Human Factors, 35*(2), 345–359.

Gawande, A. (2010). *The checklist manifesto: How to get things right*. Metropolitan Books.

Hikida, K., Matsuzaki, N., Yamamoto, S., Sakane, Y., Murata, S., Ogawa, M., Kusunoki, M., & Yoshimura, K. (2015). *The human error reduction effect of point and call checks on maritime training*. 7th International Conference on Emerging Trends in Engineering & Technology, pp. 157–159.

Reason, J. (2013). *A life in error: from little slips to big disasters*. Routledge.

Suto, I., & Ireland, J. (2021). Principles for minimizing errors in examination papers and other educational assessment instruments. *International Journal of Assessment Tools in Education, 8*(2), 310–325.

Suto, I., Williamson, J., Ireland, J., & Macinska, S. (2023). On reducing errors in assessment instruments. *Research Papers in Education, 38*(3), 357–377.

© Cambridge University Press & Assessment 2023

Appendix

Assessment Analyst checklist

1	Are all the questions answerable (i.e., candidates who have studied the full course should be able to make a sensible attempt at answering each question)?	
2	Are all the questions clear, unambiguous and without risk of confusing candidates?	
3	Is the subject content of all the questions factually accurate?*	
4	Are all the necessary visuals and resources provided (i.e., those in the question paper such as graphs, tables, images, text extracts, sources or equations, and those that are separate such as pre-release materials, inserts, etc.)?	
5	Are all the visuals and resources clear, complete, factually accurate and consistent with all other aspects of the question?	
6	Do all multiple choice and other selected response questions have the right number of response options and the right number of correct responses (e.g., none of the distractors could be interpreted as correct even if knowledge beyond the specification is used)?	
7	Does the paper avoid inconsistencies within questions (e.g., terminology, subject content)?	

*Check facts that are critical to answering the question at the appropriate level. For example, there is no need to check the accuracy of data from an original source/original research but it is important to check the attribution of sources where relevant and critical to answering the question (e.g., for history papers).

© Cambridge University Press & Assessment 2023