Revising the A2 Key and B1 Preliminary Speaking exam

Kathryn Davies
Assessment, Cambridge Assessment English

Nick Glasson
Assessment, Cambridge Assessment English

CONTENTS

Introduction

Cambridge Assessment English periodically reviews all of its assessments to guarantee fitness for purpose. The review of A2 Key and B1 Preliminary was carried out to ensure that these exams remain relevant to test users' evolving needs. Further aims were to facilitate progress up the 'proficiency ladder' through better alignment with exams at higher levels, improve alignment with the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001), and have a positive impact on teaching and learning.[1]

Cambridge English has an established process for exam revision, and for Speaking components it is summarised in Figure 1.

```
┌─────────────────────────────┐
│ Consulting stakeholders on  │
│     current exam format     │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│   Internal review of design │
│   aspects to focus on in trials │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│  Development of initial trial │
│          materials          │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│           Trial 1           │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│        Trial 1 review       │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│   Redesign of trial materials │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│           Trial 2           │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│        Trial 2 review       │
└─────────────────────────────┘
              ↓
┌─────────────────────────────┐
│     Final recommendations   │
└─────────────────────────────┘
```

Figure 1: Outline of revision process for Speaking

1. Where reference is made to 'A2 Key', this should be read as inclusive of A2 Key for Schools, and likewise the term 'B1 Preliminary' within this article encompasses the standard and the variants for schools. These exams were previously known as *Cambridge English: Key* and *Cambridge English: Preliminary*.

Consulting stakeholders

As part of its cyclical review process, Cambridge English gathered feedback from over 500 stakeholders (Teachers, Heads of English/English Co-ordinators, Directors of Studies, Centre Exams Managers and Exam Administrators) across several countries (Spain, Italy, Russia, Greece, Romania, Cyprus and Serbia) as well as from a number of assessment experts (Professional Support Leaders, Team Leaders, Chairs, Item Writers), so that both the needs of test users and the expertise of assessment specialists could inform the revision of the Speaking component of A2 Key and B1 Preliminary as effectively as possible. Feedback was collated on the basis of findings from a large-scale survey and focus groups, with key stakeholders including teachers and Centre Exams Managers.

A2 Key

The current test structure is provided in Table 1.

Table 1: Current A2 Key Speaking format

	Interaction pattern	Input
Part 1 Interview 5–6 minutes	Interlocutor asks questions to each candidate in turn, giving factual or personal information.	Spoken questions provided by the interlocutor frame.
Part 2 Information exchange 3–4 minutes	Candidates ask and answer questions about factual, non-personal information.	Spoken and written rubrics. Visual stimuli given in the candidate booklet (see Figure 2 for an example).

Consultation activities, drawing on expert opinion as well as insight from focus groups with key stakeholders, highlighted several areas of potential focus for revision trialling. In the case of A2 Key, there was a concern that the current information gap task (see Figure 2) in Part 2 did not differentiate candidates' levels enough, as evinced by one Centre Manager's comment:

'In my experience, the candidates who have shown themselves to have different levels in speaking in class often get similar results at the Speaking test.'
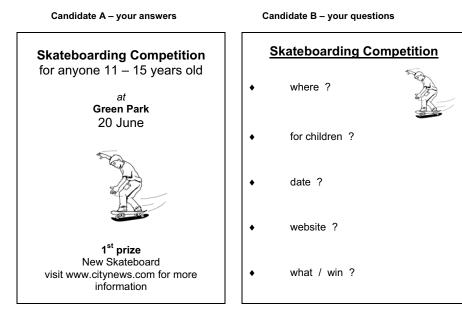
Candidate A – your answers

Skateboarding Competition
for anyone 11 – 15 years old

at
Green Park
20 June

1ˢᵗ prize
New Skateboard
visit www.citynews.com for more information

Candidate B – your questions

Skateboarding Competition

♦ where ?

♦ for children ?

♦ date ?

♦ website ?

♦ what / win ?

Figure 2: Current A2 Key Speaking Part 2 sample task

There were also indications that the current Part 2 task was not allowing more able candidates to demonstrate their full speaking performance at this level. Dissatisfaction was also reflected in survey feedback (see Figure 3).
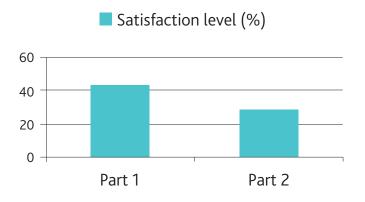
Satisfaction level (%)

A bar chart showing satisfaction levels. Part 1 is approximately 43% and Part 2 is approximately 29%. The y-axis is labelled from 0 to 60 in increments of 20.

Figure 3: A2 teacher perceptions of the current A2 Key exam format

Survey findings also revealed that developing learners' speaking ability at A2 level was a major concern for teachers. As one teacher commented, their main challenge was 'to encourage students to speak'; another said that 'the greatest challenge is to make them fluent in speaking and taking turns'. Feedback of this kind suggested, at least in part, that the washback effect of the current exam format was not conducive to building learners' interactive skills, for example simple turn taking.

Greater alignment with Level A2 of the CEFR so that candidates have the opportunity to fully demonstrate their ability across a broader range of speaking sub-skills and language functions was thus of fundamental importance in the revision.

According to the CEFR (Council of Europe 2001:58), candidates who have reached A2 level can:

- give a simple description of people, living (and working) conditions and daily routines as a short series of simple phrases and sentences
- explain what they like or dislike about something
- show their ability to manage simple and routine exchanges of ideas and information on familiar topics, provided the other person helps if necessary.

In line with survey feedback, the new test design would aim to provide better measurement and better support for teachers as they prepare their learners for A2 Key Speaking (positive washback).

B1 Preliminary

Table 2: Current B1 Preliminary Speaking format

	Interaction pattern	Input
Part 1 Interview 2–3 minutes	Interlocutor asks candidates questions to elicit personal information.	Spoken questions provided by the interlocutor frame.
Part 2 Collaborative task 2–3 minutes	Interlocutor delegates a collaborative task (simulated situation) to the pair of candidates.	Spoken rubrics. Visual stimuli given in the candidate booklet (line drawings).
Part 3 Long turn 3 minutes	Interlocutor delegates an individual task to each candidate.	Spoken rubrics. Visual stimuli given in the candidate booklet (one photograph per candidate).
Part 4 Discussion 3 minutes	Interlocutor initiates a discussion between the candidates.	Discussion set up by interlocutor using interlocutor frame.

In the case of B1 Preliminary, comments from experts and stakeholders during focus groups tended to concern task order rather than task design. Of particular importance were expert appraisals of how effectively the current Part 3 (collaborative task) and Part 4 (extended discussion) tasks matched their original aims as a consequence of the order in which they appear. Part 4 is intended to be an interactive task following on from the 'long turn' but its reliance on the Part 3 content had the effect of limiting how generative it can be. As one assessment specialist with extensive examining experience noted, Part 3 is always constrained by the photos preceding it; this can result in questions which do not generate very much language, especially for stronger candidates.

There was also a concern about the level of agency assumed by the Part 4 task in that candidates were more or less left to manage the interaction entirely on their own. This was something commented on by one external assessment specialist:

> 'Currently the aim in Part 4 is for candidates to interact with no support (examiners giving this only if necessary). Part 4 in higher-level tests (B2 First, C1 Advanced and C2 Proficiency) is conducted by the examiner, but allowing for candidates to develop their answers. It could be argued that B1 level candidates actually need more support than the higher levels.'

Comments from several experienced examiners suggested that the final interactive element would more logically follow from the Part 2 discussion task rather than the long turn.

Across both A2 and B1 exams, there was an additional concern to create greater 'family resemblance' across the exam suite as a whole so that from A2 to C2 there is consistency in exam structure as far as is feasible. This enables reduction of test anxiety among learners and also supports teachers preparing students for the exam by standardising the test format.

Having gathered extensive feedback from all relevant sources, assessment specialists focused on developing initial trial test specifications. This information was used to judge how to improve measurement of the construct. For example, in the case of A2 Key, the new task assesses candidates' ability to 'participate in short conversations in routine contexts on topics of interest' (Council of Europe 2017:85).

The revision of assessment scales is typically a much broader and lengthier activity involving all CEFR levels and all assessments aligned to the scales. (Galaczi, ffrench, Hubbard and Green 2011 outline the Cambridge English approach to such work.) However, the existing assessment criteria were considered to accurately represent the A2 and B1 constructs for Speaking and therefore were not changed.

Trial 1: Tasks used

A2 Key

The revised test structure is provided in Table 3.

Table 3: Revised A2 Key Speaking format

	Interaction pattern	Input
Part 1 Interview 3–4 minutes	Interlocutor asks questions to each candidate in turn, giving factual or personal information.	Spoken questions provided by the interlocutor frame.
Part 2 Discussion task 5–6 minutes	Candidates discuss likes, dislikes and give reasons.	Spoken and written rubrics. Visual stimuli given in the candidate booklet (see Figure 4 for an example).

The format for Part 1 remained the same: an interlocutor-led interview. In Phase 1, the interlocutor asks the candidates questions, in turn, using a standardised script to ensure all candidates have the same opportunity to perform. Candidates give basic personal information of a factual nature.

In Phase 2, candidates respond to personal questions, in turn, on two familiar topics such as daily life, school, leisure activities, family, etc. The first two questions require brief answers only. Each candidate is also asked to give an extended response to a prompt connected to the first two questions ('please tell me something about…'). Previously, interlocutors were given the freedom to decide which questions to address to which candidate(s), and questions were chosen at random across a number of topics.

The revised task provides greater coherence as questions relate to just two topics. The more prescriptive frame also supports examiners in their ability to be consistent thereby ensuring equal opportunity for candidates.

The new Part 2 task takes the form of a collaborative discussion. It provides greater opportunities for candidates to fully demonstrate their speaking ability and their interactive communication skills through a more personalised, authentic and meaningful exchange.

In Part 2 Phase 1 candidates are invited to talk together about a topic. They are provided with visual stimuli and asked if they like the activities, places or things depicted and to say why or why not (see Figure 4).

Do you like these different holidays?

Figure 4: Revised A2 Key Speaking Part 2 sample task

The interlocutor is allowed up to a maximum of two minutes for candidates to talk together independently, before coming in with prompts aimed at extending the discussion and encouraging candidates to develop their utterances, for example, 'Do you think camping is fun?', 'Why (not)?' Interlocutors end this part of the exam with a closing question directed at each candidate in turn; in the case of the sample task: 'Which of these different holidays do you like best?'

Part 2 Phase 1 aims to assess candidates' ability to interact with a partner and with the interlocutor, to express likes and dislikes, and to give simple reasons. Candidates are invited to express opinions about the different activities, things or places represented but are expected to talk about these only in relation to themselves and their experiences of the world, as is appropriate for A2 level.

In Part 2 Phase 2, the interlocutor asks each candidate two more questions, broadening the topic of Phase 1. Phase 2 aims to assess candidates' ability to talk more freely on the topic discussed in Phase 1. Candidates are given the opportunity to demonstrate their full speaking ability in a less formulaic but still supported manner in this last part of the test.

B1 Preliminary

For B1 Preliminary the focus of initial trialling was the following:

● the re-ordering of the tasks, so that the photo-based individual turn task occurs before the discussion task

● the removal of the follow-on question phase from the photo-based individual turn tasks and the inclusion of a follow-up question phase after the Part 3 discussion task (see Figure 5)

● the use of a 'split' rubric in the discussion task (see Figure 6), similar to that of B2 First Speaking Part 3.

Figure 5: Part 4 follow-on questions (revised format)

Figure 6: Example of split rubric interlocutor frame (Part 3)

Tasks were created to be trialled on both pairs and groups of three.

Trialling cohorts

A2 Key

Eight Speaking Examiners participated in initial trialling, each with at least six years' experience of A2 Key Speaking exams, and, as a group, covering a diverse candidature across Brazil, Czech Republic, Italy, Spain and the UK.

In the trialling, Examiners were invited to watch videos of candidates taking the revised exams and to rate their performances. They were also asked to provide feedback on the new exam by completing a questionnaire.

Candidates were deemed by the Examiners to be typical for the ability level targeted by the exam and to have produced sufficient language to allow Speaking Examiners to rate them across all three assessment criteria.

B1 Preliminary

There was a mix of abilities in this initial trialling cohort but most were students intending to take B1 Preliminary.

The UK-based trialling provided a very diverse range of students from Iran, Albania, Korea, France, Morocco, Saudi Arabia, Japan, Libya, China, Thailand, Colombia, Brazil, Turkey, Czech Republic and Armenia.

Trials overseas featured monolingual pairs, as one might expect, and the sample was intentionally limited in each location in order to provide a representative sample of a diverse range of language groups overall. Despite this, the sample did cover a range of abilities at the level in every case – from strong to weak. Trials were carried out by very experienced Speaking Examiners.

This first phase of trials involved trialling of full tests on over 60 candidates. In all trials, a current-format test was also administered to provide an insight into how the proposed and current designs compared.

Observations and feedback were gathered via standardised forms and following a set of trialling instructions (an excerpt of a trialling observation form is provided in the Appendix). Trials were filmed and in August 2016 a small-scale marking and examiner survey was also carried out using these video recordings.

Following trials, an internal review was conducted where feedback and outcomes were considered.

Trial 1: Review

A2 Key

Overall, Examiners' feedback was positive. The revised tasks were considered to be an improvement on the current format.

Examiners reported that:

- the new tasks provide greater and richer opportunities to assess candidates' interactive communication skills
- the new tasks elicit more real-life realistic language and interaction types than the current ones
- the new tasks are in line with other Cambridge English Qualifications.

Some illustrative comments from the Examiners:

> 'The previous Part 2 was quite "scripted" and candidates produced a narrower range of language. This task allows the candidates to interact in a meaningful way, and produce language which is their own, rather than relying on the previous prompt questions/information card, which often was misinterpreted or relied on candidates being able to read out loud accurately.'

> 'It fits better with classroom practice and is less rehearsed.'

> 'I think the new test format, particularly Part 2, is a better test of language and interactive communication.'

> 'It's a huge improvement! So much more suitable for this level of candidates and elicits a much wider range of language than the current version.'

Overall, the difficulty level of the new tasks was judged to be appropriate by Examiners, i.e. at A2 level and similar to the difficulty level of the current format. All Examiners agreed that the revised exam allowed stronger candidates to show their speaking ability beyond A2 level. The new format of Part 2 in particular gives candidates greater autonomy, thus enabling them to fully show their ability. The flipside of this was a concern that this could be perceived as a reduction in the support provided to weaker candidates. Indeed, there was a sense in the trialling that the new exam might suit stronger candidates better.

Three out of eight examiners felt that the instructions in Part 2 lacked clarity and that some candidates did not understand the task requirements, leading them to describe the pictures rather than react to them with personal opinions. This lack of clarity could have contributed to concerns around the suitability of the task for weaker candidates.

Finally, feedback from examiners revealed that the anticipated discussion between candidates in Part 2 was not always in evidence as candidates did not respond to each other's utterances. Assessment of candidates' interactive communication was therefore based entirely on their responses to the interlocutor's questions.

Some illustrative comments from the Examiners:

> 'It wasn't clear that the candidates understood what to talk about with the pictures.'

> 'Some of them described the pictures in Part 2 rather than giving their opinion of the activities/places, but using the prompts seemed to stop them doing this.'

> '…[V]ery few of these trial candidates asked each other a question in the discussion phase – the children in particular just talked about themselves, without really linking what they said to their partner's contribution.'

> 'When asked to "tell each other" and talk together, some candidates dried up until prompted further.'

These findings fed into recommendations for subsequent trialling. The concept for the new Part 2 task had been proven, but the execution required further refinement; specifically we sought to improve the clarity of the rubric.

B1 Preliminary

The first phase of trials indicated that while the proposed task order was seen as positive based on expert appraisal and evidence from trial footage, the split rubric in the discussion task was problematic in a number of ways for B1 level:

- It often led to repetition, as candidates reached a decision prior to the second rubric being delivered by the interlocutor.

- In the context of this B1 exam, it seemed inauthentic and artificial to divide the appraisal of a range of (relatively concrete) options with the decision of 'which is best', etc.

- At Cambridge English we typically draw a distinction between Speaking at B1 being focused on 'negotiating agreement' and Speaking at B2 stressing 'reaching a decision through negotiation', due in part to the more concrete operations expected of B1 learners.

- From a task-writing perspective, the removal of the focus of the discussion had the effect of making the task scenario inadequate as a springboard for a developed discussion when compared to the current task design. It was hard to 'move the discussion on' at B1.

It was noted in trials that the split rubric was problematic at B1 level:

> 'Having trialled the split rubric task in isolation from other suggested changes, we were very much of the opinion that the task format wasn't working at this level. The splitting of the task into two phases seems to be artificial. Working towards a conclusion and discussing the relative merits and demerits of the various options is very much one operation in the current test.'

> 'The removal of the intended outcome of the discussion rendered the context rather thin and made the first part of the task rather abstract. Strong candidates tended to fill this vacuum with an imagined outcome of their own, whilst weaker candidates floundered in a rather abstract discussion that was leading nowhere.'

It could be argued that the demands of a more abstract discussion will generally tend to favour those with greater interactional competence as a result and this was borne out in trials for the split rubric here, where the stronger candidates managed the task better.

In her study of interactional competence, this difference in capabilities is something Galaczi (2014) observes: 'the interactional profile of B1 learners was found to be generally characterised by low mutuality between the speakers' (Galaczi 2014:560). By contrast the B2 learners' better-developed linguistic resources and automaticity in processing allow them to be both focused on constructing their own response and decoding their partner's contributions: 'B2 test takers were found to be more adept at keeping active the roles of speaker and listener at the same time' (Galaczi 2014:564).

Trial 2: Tasks used

A2 Key

The wording of the rubric and the timing allotted in Part 2 Phase 1 were amended in an attempt to provide greater clarity and support, thereby responding to feedback received during initial trialling. A second stage of trialling was subsequently undertaken.

The Part 2 Phase 1 rubric, 'Tell each other what you think about…' (see Figure 7), was replaced with a direct question – 'Do you like…?' (see Figure 8). By reducing the structural complexity of the instruction, the processing load was lowered and task requirements, it was anticipated, would be clearer and easier to grasp.

Tell each other what you think about these different holidays.

I'll say that again.

Here are some pictures that show different holidays.
Tell each other what you think about these different holidays.

OK? Talk together.

Figure 7: Trial 1 Part 2 Phase 1 rubric

> Do you like these different holidays? Say why or why not. I'll say that again.
>
> Do you like these different holidays? Say why or why not.
>
> All right? Now, talk together.

Figure 8: Trial 2 Part 2 Phase 1 rubric

To ensure adequate support for weaker candidates, the timing requirements of Part 2 Phase 1 were also revised. In Trial 1, interlocutors were instructed to 'allow up to two minutes for discussion'. In Trial 2 this was reduced to 'allow up to one minute', thus allowing the Examiner to intervene and lend support to candidates earlier.

This was later revised to 'a minimum of one minute and a maximum of two' to allow Examiners the flexibility to tailor timing requirements to the needs of candidates. While some candidates at this level could only sustain interaction without help from the interlocutor for one minute, some candidates appeared able to sustain it for longer. In the case of stronger candidates, or candidates who took time to warm up, interjecting after one minute risked interrupting the flow and not giving candidates the opportunity to extend their discussions fully.

B1 Preliminary

A summary of the revised test structure is provided in Table 4.

Table 4: Revised B1 Preliminary Speaking format

	Interaction pattern	Input
Part 1 Interview 2–3 minutes	Interlocutor asks candidates questions to elicit personal information.	Spoken questions provided by the interlocutor frame.
Part 2 Long turn 3 minutes	Interlocutor delegates an individual task to each candidate.	Spoken rubrics. Visual stimuli given in the candidate booklet (one photograph per candidate).
Part 3 Collaborative task 2–3 minutes	Interlocutor delegates a collaborative task (simulated situation) to the pair of candidates.	Spoken rubrics. Visual stimuli given in the candidate booklet (line drawings).
Part 4 Discussion 3 minutes	Interlocutor leads a discussion with candidates.	Spoken questions provided by the interlocutor frame.

The 'split' rubric was not included in the second stage of trialling for the reasons outlined in the section on Trial 1. Instead, a version of the current task rubric was developed which avoided the repetition of the rubric (see Figure 9).

Part 3	
Interlocutor	Now, in this part of the test you're going to talk about something together for about two minutes. I'm going to describe a situation to you. *Place **Part 3** booklet, open at **Task 1**, in front of the candidates.* **A young man works very hard, and has only one free day a week. He wants to find an activity to help him relax.** **Here are some activities that could help him relax.** **Talk together about the different activities he could do, and say which would be most relaxing.** All right? Now, talk together.
Candidates 🕐*approx. 2–3 minutes*	....
Interlocutor	Thank you. (Can I have the booklet please?) *Retrieve **Part 3** booklet.*

Figure 9: Example of Phase 2 Part 3 rubric (non-split)

Another significant change in Phase 2 was the uncoupling of the photos used in the individual turn tasks. Previously these were linked thematically (e.g. 'A day out') but this was felt to potentially advantage or disadvantage candidates, while also making the successful development of these tasks far more difficult.

For Phase 2, the photographs used were deliberately paired so that they would *not* overlap in basic topics or themes. There were also back-up prompts (see Figure 10) added to the interlocutor frame for the individual turn to provide additional means for interlocutor support.

Back-up prompts
- Talk about the people/person.
- Talk about the place.
- Talk about other things in the photograph.

Figure 10: Back-up prompting in Part 2 (individual turn)

There were also more minor alterations to Part 1, based on observations from the first phase of trials (e.g. slight changes to the introductory rubrics to ensure a better flow of questions).

Trialling cohorts

A2 Key

Seven Senior Examiners administered the revised tasks to 127 candidates across six countries (Argentina, Greece, Romania, Russia, Taiwan and the UK). Qualitative analysis of their feedback was conducted. Seventeen Russian candidates sat the pre-revision and the revised Part 2 tasks so that a direct comparison of candidate performance across the two formats could be made. This was completed via functional analysis of candidates' speech, achieved by comparing the number of language functions elicited by both formats (see Figure 11).

Candidates were deemed by the Examiners to be representative of the target candidature for the test and to have produced sufficient language to allow Speaking Examiners to rate them across all three assessment criteria.
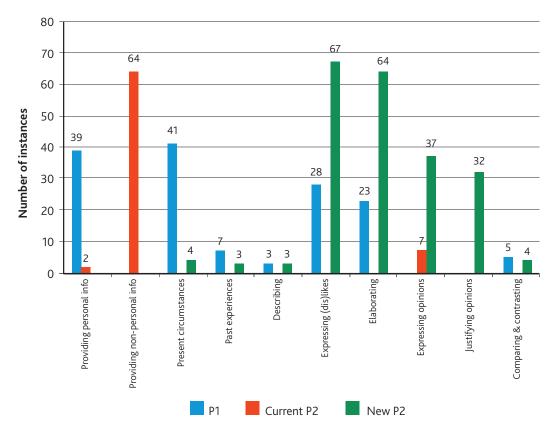
Figure 11: Functional analysis of language in trialling of A2 Key Speaking (P = Part)

B1 Preliminary

Phase 2 trials were carried out in a similar wide range of locations and with an equally wide range of ability levels as in Phase 1. As in the first phase, trial materials were complete tests and these were carried out alongside a current-format test to provide comparison data and feedback. More than 80 candidates were involved in this phase in locations in Russia, Argentina, Taiwan, the UK, Italy, Romania, Vietnam and Greece. All samples were recorded for analysis, examiner surveys and feedback activities.

Trial 2: Review

A2 Key

Examiners were consistent in their opinions that the revised exam, and especially Part 2, was an improvement on the pre-revised task. Themes that stood out from their feedback are as follows:

- it allows candidates greater flexibility and more opportunities to demonstrate their A2-level knowledge and speaking skills, particularly their interactive skills, more fully
- it allows especially strong candidates to demonstrate their speaking skills beyond the requirements of A2 level
- it elicits more personalised language: candidates can produce novel utterances
- it elicits a wider range of language functions, e.g. responding to opinions and turn-taking

- it focuses more on meaning than on form, thereby indicating its potential for positive washback, i.e. a focus on communicative language use rather than formal accuracy

- it allows for more meaningful and more authentic interaction

- it increases candidate enjoyment and overall performance

- it is in line with other Cambridge English Qualifications.

Some illustrative comments from the Examiners:

> 'I totally like this new format and I think it can give candidates better opportunities for speaking and using more language. Pictures give candidates more independence.'

> 'The difference [in the new Part 2 versus current] in the quality of students' utterances and their resultant performances was stark … students consistently performed so much better in the revised Part 2.'

> 'Students responded very well … and participated in lively discussions. Even though they were not prepared for a task like this, they managed to sustain a simple discussion.'

> 'One of the students said: "I love to talk about things with my friends. The other thing we did was not so interesting."'

B1 Preliminary

Extensive feedback was taken from Examiners involved in the second phase of trials. The vast majority of this feedback endorsed the new proposed test design, which was felt to 'flow' more naturally and allow the candidate time to warm up via the individual turn prior to the collaborative discussion task.

No evidence from trialling suggested that the use of different topics/themes in the photo-based task would disadvantage either candidate. The use of different topics/themes also limits the potential for candidates to 'lift' language from each other.

In Part 3, the revised discussion task rubric indicated in trials that it worked well and that the removal of the repetition of the main rubric was not impacting on the candidates' ability to perform the task.

The use of a follow-on set of questions in Part 4, after the discussion task, was also felt to be a positive move as it meant B1 candidates were no longer required to take on an interlocutor-like role in the interaction and the examiner was more able to re-balance the contributions from candidates, as in B2 First, B2 First for Schools and C1 Advanced Part 4.

This was seen as preferable to the current B1 Preliminary Part 4 task, which often elicited two further 'long turns' from candidates rather than a genuine interaction. This also meant the examiner could step in if candidates 'dry up' in their response while also giving scope for some further interaction (i.e. by the interlocutor directing a question to both candidates). The new Part 4 still afforded the assessors scope to fine-tune their marks in the final phase of the test event too.

Conclusions

A2 Key

In conclusion, the new exam is considered to be an improvement on the current format. It elicits a wider range of language and language functions, thus allowing candidates to fully demonstrate their speaking skills and providing a more authentic and meaningful task. Despite the resounding positive appraisal of the new tasks, Cambridge English remains mindful of the need for clarity of instruction and expectations for those who sit the exams as well as for those who administer them. As part of the rollout of the new format, we will ensure that we:

- provide information on the focus of Part 2 to candidates and their teachers, highlighting that the pictures are intended to prompt discussions about the activities, places and things represented and that the task is not to describe them, and that candidates should be encouraged to respond to their partner's utterances

- include advice and appropriate back-up questions for Speaking Examiners to help guide candidates through the task, to provide appropriate scaffolding and support, and to allow candidates to demonstrate their speaking skills fully.

B1 Preliminary

In conclusion, it was felt the revised exam format for B1 Preliminary was one that provided much greater interlocutor control than the existing test design, and improved the test experience for candidates without diminishing the test's ability to make accurate assessments of candidates at this level. The focus on the CEFR B1 level is maintained, but the revised test also allows stronger candidates to show a fuller range of skills and also aims to support less able candidates more than previously.

References

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.

Council of Europe (2017) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with New Descriptors*, available online: rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Galaczi, E (2014) Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics* 35 (5), 553–574.

Galaczi, E, ffrench, A, Hubbard, C and Green, A (2011) Developing assessment scales for large-scale speaking tests: A multiple-method approach, *Assessment in Education: Principles, Policy & Practice* 18 (3), 217–237.

Appendix: Excerpt from observation form used in trialling

For quantitative observations of language functions, the following type of form was used:

Informational functions/features	Part 1	Part 2	Part 3	Part 4
Providing personal information				
Providing non-personal information (e.g. dates/times/prices)				
Talking about present circumstances				
Talking about past experiences				
Talking about future plans				
Describing people, places and situations				
Expressing preferences				
Expressing opinions				
Justifying opinions				
Elaborating (e.g. explaining and giving reasons)				
Comparing and contrasting				
Suggesting and recommending				
Any other functions? [Please state below]				

Informational functions/features	Part 1	Part 2	Part 3	Part 4
Agreeing/disagreeing				
Asking for opinion				
Asking for information				
Negotiating meaning: - checking meaning - asking for clarification - correcting utterance(s)				
Responding to required clarification				
Paraphrasing words and ideas if necessary				
Any other functions? [Please state below]				

Informational functions/features	Part 1	Part 2	Part 3	Part 4
Responding appropriately				
Initiating				
Turn-taking				
Sustaining a long turn				
Any other functions? [Please state below]				

The following examples of trial feedback questions are taken from a standardised form focused on more qualitative judgements on trialling of the 'split' rubric in B1 Preliminary Speaking (Phase 1):

How would you describe the candidates' response to the 'split rubric'?
For example:

- Was the transition from one phase of the task to the other smooth?

- Did any of the rubrics require repetition or clarification?

- As an Examiner, did you feel the rubrics were easy to read/manage?

- Were the rubrics clearly understood?

- Was there overlap or repetition in language produced between the discussion and decision phases?

- How did the suggested timing for the candidate response compare to reality?

How would you describe the candidates' response to the Part 4 task?
For example:

- As an Examiner, did you feel the rubrics were easier to read/manage?

- In comparison with the current *Cambridge English: Preliminary* and *Cambridge English: Preliminary for Schools* Part 4 task, did you feel this was better in terms of managing the test experience for candidates?

- In comparison with the current *Cambridge English: Preliminary* and *Cambridge English: Preliminary for Schools* Part 4 task, did you feel this was better in terms of providing a good sample of language (particularly re: 'fine-tuning' of marks)?

- In comparison with the current *Cambridge English: Preliminary* and *Cambridge English: Preliminary for Schools* Part 4 task, did you feel this was better in terms of providing an equal contribution from both candidates?

- How did the suggested timing compare with the real time taken?

- Were all the questions you used clear and understood?

- How many questions did you use in total? Did you make much use of the additional prompts (e.g. 'What do you think?')?