



C2

C1

B2

B1

A2

A1

Pre A1

Cambridge English Qualifications

Pre A1 Starters, A1 Movers and A2 Flyers

The 2018 Pre A1 Starters, A1 Movers and A2 Flyers revisions



Contents

Overview	3
The 2018 Pre A1 Starters, A1 Movers and A2 Flyers revisions	
Chapter 1	7
Developing new Pre A1 Starters, A1 Movers and A2 Flyers speaking assessment scales	
Chapter 2	14
Creating an easy-to-understand alignment for Pre A1 Starters, A1 Movers, A2 Flyers, A2 Key for Schools and the CEFR	
Chapter 3	18
New writing tasks for young learners: A1 Movers and A2 Flyers	
Chapter 4	25
Revising tasks to ensure children have the most opportunity to demonstrate their English skills	
Chapter 5	28
Updating the Pre A1 Starters, A1 Movers and A2 Flyers vocabulary lists	
Appendix	31
References and further reading	33
Authors	34

Overview

The 2018 Pre A1 Starters, A1 Movers and A2 Flyers revisions

Sarah Albrecht, Maggie Dunlop

REASONS FOR THE REVISIONS PROJECT

The Pre A1 Starters, A1 Movers and A2 Flyers exams are a set of three English tests specially designed for primary school-aged children (approximately ages 6 to 12) who are in the early stages of English language acquisition. Pre A1 Starters covers the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) pre-A1 level. A1 Movers covers CEFR A1 and pre-A1 levels. A2 Flyers covers CEFR A2, A1 and the top end of pre-A1 levels. Each exam has three papers: a Listening paper, a combined Reading/Writing paper and a Speaking test. The exams are part of the Cambridge English Qualifications. Prior to October 2017 the exams were known as *Cambridge English: Starters, Movers and Flyers*.

The exams, like all Cambridge English Qualifications, are designed according to the Cambridge Assessment English principles of good practice (Cambridge English Language Assessment 2016). These principles are summarised in Figure 1 below.

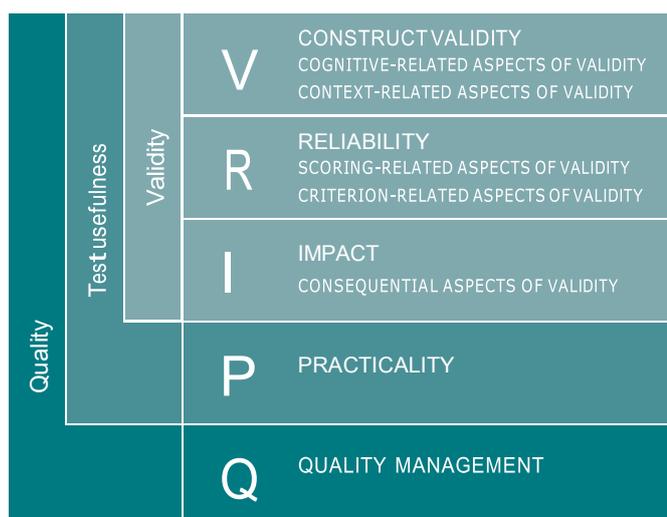


Figure 1: The Cambridge Assessment English VRIPQ approach to implementing the principles of good practice

Regarding construct validity claims, the exams are designed to assess children's English language use and development, so the tasks are carefully designed to be fun and age appropriate in terms of topic, genres and cognition requirements. Impact is another area of high importance in assessment of children, as children's attitudes towards English, school and learning are in formative stages. Therefore the Pre A1 Starters, A1 Movers and A2 Flyers exams are designed to be enjoyable, confidence-enhancing activities for children that boost their enthusiasm for English learning. A key feature of this design is a purposeful 'no pass, no fail' design, in which all children receive a certificate and receive at least one shield (out of a total of five for each paper).

As part of the Cambridge Assessment English principles of good practice, there is an ongoing commitment to maintaining the quality of Cambridge English Qualifications. Therefore all exams are regularly reviewed to ensure they remain relevant to learners and schools. The Pre A1 Starters, A1 Movers and A2 Flyers revision project started in 2014, with the revised exams launched in January 2018. The main objectives of the revision were:

- to ensure that the exams continue to accurately reflect children's achievement in English
- to ensure that Pre A1 Starters, A1 Movers and A2 Flyers continue to meet the evolving needs of candidates, their parents and teachers.

At the same time, it was important to ensure that the exams continue to offer children a positive, confidence-boosting test experience that motivates them to continue learning English.

DECIDING WHAT TO REVISE

When conducting research and revision activities, Cambridge Assessment English typically adopts a mixed methods approach to triangulate the multiple types of data that can be derived from diverse sources (Moeller, Creswell and Saville (Eds) 2016). In this revision project, data sources included teacher, candidate and parent questionnaire and interview feedback, examiner and test developer comments, task and paper trialling with subsequent quantitative inferential analyses, task and response content and linguistic analysis, and expert review. This range of data sources provided answers to the research questions posed at each stage of the revision project.

The very first question asked in the revision project was: What needs changing, and what should stay the same? To answer this question, data from a number of sources was analysed. These sources included surveys of key stakeholders, internal and external expert reviews, and psychometric analysis of tasks.

The survey of key stakeholders showed very high satisfaction with all tasks at Pre A1 Starters, A1 Movers and A2 Flyers, and the same very high satisfaction with the appropriateness and variety of topics, along with the look and feel of the exams at all three levels. However, there were clear requests for clarification on what the shields meant. For example, five shields on A1 Movers Listening was the same as how many shields on A2 Flyers? How do A2 Flyers shields match up with an A2 Key for Schools 'pass'? And how many shields on A1 Movers were needed to achieve CEFR A1 proficiency? In addition, many stakeholders expressed a desire for more information, to serve instructional purposes.

Regarding the reports from experts in language assessment, as well as the psychometric analyses of tasks, the construct of children's emerging English language skills was found to be well represented by the test tasks. In particular, the test tasks were found to effectively encourage children to use their English skills. The tasks were found to achieve this by providing engaging, age-appropriate activities and images, and by providing age- and level-appropriate support.

That said, some areas for improvement were observed. In particular, more comprehensive coverage of children's emerging English writing skills was identified as a priority. In addition, several minor possible improvements to individual test tasks were identified, with the purpose of making it easier for candidates to show their English skills. Finally, the revision project was also recognised as an opportunity to update the Young Learners vocabulary list.

As a result of the various studies, several sub-projects were therefore carried out:

- An alignment project was conducted to make clear:
 - how shields on Pre A1 Starters, A1 Movers, and A2 Flyers align with each other
 - how A2 Flyers and A2 Key for Schools align with each other
 - how Young Learners exams align with the CEFR.
- New level-appropriate writing tasks were developed for A1 Movers and A2 Flyers to provide better coverage of children's emerging writing skills.
- Minor changes were made to a selection of tasks, to make it easier for children to show their English skills.
- The vocabulary list was updated to reflect children's current usage of English.

WHAT REVISIONS WERE MADE

The alignment project was a significant piece of work consisting of several components, including:

- developing new speaking assessment scales that are aligned with each other and the CEFR, and that provide enough precision to identify achievement across CEFR pre-A1, A1 and A2
- ensuring task types across Pre A1 Starters, A1 Movers and A2 Flyers are consistent where possible, to improve content alignment between the exams
- conducting psychometric scale alignment work to ensure Pre A1 Starters, A1 Movers and A2 Flyers shields are clearly matched to each other, to A2 Key for Schools and to the CEFR.

The new speaking assessment scales (see Chapter 1) have provided a clear progression in speaking assessment criteria from Pre A1 Starters to A1 Movers to A2 Flyers, and added precision to the scoring of children's spoken English proficiency. Both these new features were necessary to align the Pre A1 Starters, A1 Movers and A2 Flyers Speaking tests with each other, with A2 Key for Schools, and with the CEFR.

In addition, minor modifications to the Listening and Reading/Writing paper tasks were made to align the exams more consistently with one another (see Chapter 2). The focus of the tasks across the three levels is now more consistent and the step up between each level is now clearer.

Finally, using the revised papers and assessment scales, a psychometric alignment study was conducted to ensure all post-revision exam versions are consistently aligned (Chapter 2). The Rasch-based scales of the Pre A1 Starters, A1 Movers and A2 Flyers exams were first aligned to each other and to the A2 Key for Schools exam. This created an alignment between the three levels and the CEFR, from which a clear alignment between shield scores on different levels and to the CEFR was developed. Finally, a system of an unbroken chain of internal anchors was set up, allowing for exams to be pretested and calibrated and later refined on full live data.

To address the need for writing tasks on the A1 Movers and A2 Flyers exams, new level-appropriate 'extended' writing tasks have been introduced to A1 Movers and A2 Flyers (see Chapter 3), replacing the pre-revision Part 2 on each exam that was identified for possible removal. Adding these new tasks ensured that candidates' ability to communicate in writing is more fully assessed, and the task formats adopted ensure this assessment is done in age-appropriate and level-appropriate ways. For example, the gradual reduction of support in the A1 Movers task is designed to assist less confident candidates in showing what they can do. Likewise, the increasing freedom is intended to build candidates' confidence in their ability to perform the task. Similarly, candidates' emerging ability to write extended text is tested in the new A2 Flyers story-writing task, as befits the level. The task format, which is story based and designed for primary-aged children, is designed to elicit interest and engagement from candidates.

During the revisions, minor modifications were also made to selected tasks on the Listening and Reading/Writing papers to ensure that children had the best opportunity to demonstrate their English skills (see Chapter 4). These modifications included removing the requirement to draw a picture on A1 Movers and A2 Flyers Listening Part 5, to avoid distracting candidates with drawing, using plural nouns in Pre A1 Starters Reading/Writing Part 1 to represent natural English usage, and changing the genre on Pre A1 Starters Reading/Writing Part 4 from a riddle to a factual text, a genre with which children are more familiar.

Regarding the Speaking tests, minor modifications were made to selected tasks to allow candidates to demonstrate their ability more fully (see Chapter 4). All candidates are now asked their name, and A1 Movers and A2 Flyers candidates also say their age, to create a comfortable environment. Pre A1 Starters candidates are now asked a 'Tell me ...' question to offer opportunity to say more, if they are able. The A1 Movers and A2 Flyers story tasks now include story titles and character names, so candidates can focus on telling the story.

Finally, the vocabulary list was updated to ensure that the words on the list remain current and relevant (see Chapter 5). Changes were based on consultation with experienced teachers of young learners, and on research involving the creation and analysis of a corpus of Pre A1 Starters, A1 Movers and A2 Flyers words. The size of the vocabulary list increased by 9% at Pre A1 Starters, 13% at A1 Movers and 12% at A2 Flyers level. Words were added if they had recently become high-frequency words (e.g. *laptop*), to complete existing lexical sets (e.g. *shorts*

for clothing), and to reflect words that were being used naturally by primary-aged English language learners (e.g. *by myself* for A2 Flyers). Words were removed when they were obsolete or becoming archaic (e.g. *policeman/woman* replaced by *police officer*), or when identified as lacking orthographic transparency (e.g. *John*) or as culture specific (e.g. *supper*).

CONCLUSION

In summary, the changes described in this report represent principled choices designed specifically to ensure that the Pre A1 Starters, A1 Movers and A2 Flyers exams remain fit for purpose. That is, that the exams continue to accurately reflect children's achievement in English, and that they continue to offer children a positive, confidence-boosting exam experience that motivates them to continue learning English, and that they support teachers in teaching practical English skills.

Chapter 1

Developing new Pre A1 Starters, A1 Movers and A2 Flyers speaking assessment scales

Maggie Dunlop, Chris Cullen

INTRODUCTION

To support the alignment project, the development of new speaking assessment scales was identified as a priority. Revisions to the speaking assessment scales primarily focused on ensuring that speaking scores provided the information required to create a clear, 'stepped' alignment between Pre A1 Starters, A1 Movers, A2 Flyers and A2 Key for Schools, although the revisions were also an opportunity to review the speaking construct the exams were assessing.

This development was required because with the previous Pre A1 Starters, A1 Movers and A2 Flyers assessment scales, although assessing over Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) pre-A1 to A2 levels, five shields could be awarded to children across a relatively wide band of proficiency. This wide range would have made it difficult to create a clear, 'stepped' alignment.

Therefore, in order to enable clear, standardised alignment between exams and with the CEFR, new scales were required so that examiners could assess English speaking proficiency with more accuracy. This more precise data provided by the examiners then enabled the alignment work that is discussed in Chapter 2.

This chapter provides an overview of the activities associated with the development of the new assessment scales. Note that this chapter describes the process by which the new speaking assessment scales were developed and validated. The work described in this chapter was to allow examiners to be more precise, and was not related to changing the number of shields that a child's performance would achieve. See Chapter 2 to understand more about how the shields relate to children's performance.

OVERVIEW OF DEVELOPMENT AND VALIDATION METHODS

Several research questions framed the development and validation process. The questions were: How can primary-age children's beginner spoken English language proficiency be described step by step? What are the key criteria by which these descriptions can be grouped? Can examiners use the new assessment scales consistently? Do the new assessment scales improve the precision of the assessment?

To answer the first and second questions, a literature review of research on development of children's beginner second language speaking proficiency was conducted. After that, first drafts of the new scales were developed and underwent external review by assessment experts. Based on this external review, revisions were made and the scales were passed to the next stage of the development and validation process.

The scales were then tested as to whether they sufficiently answered the first two questions through the second stage of the development and validation process. The second stage was an empirical review of the assessment criteria that consisted of two phases. First, 28 practitioners (experienced teachers of young learners who were also experienced Pre A1 Starters, A1 Movers and A2 Flyers examiners) reviewed the scales and provided feedback. Next, 23 experienced practitioners were given some of the new descriptors found on the scale, plus some descriptors from A2 Key for Schools, and asked to identify which test the descriptor should be for. The results of these studies were analysed (discussed below) and further revisions were made to the scales.

Finally, to find answers to the third and fourth questions, a trial was conducted in which experienced Pre A1 Starters, A1 Movers and A2 Flyers examiners marked children's speaking performances on Pre A1 Starters, A1 Movers and A2 Flyers using the old and the new assessment scales. The purpose of this study was to ensure that examiners used the assessment scales consistently, and to investigate how the distribution of scores improved from the old scales to the new scales.

INITIAL SCALE DEVELOPMENT

To start scale development, several sources were reviewed: Field's (forthcoming 2017) review of the cognitive validity of the Pre A1 Starters, A1 Movers and A2 Flyers Speaking tests, existing scales of speaking performance for young learners, documentation about the intended Speaking test construct for Pre A1 Starters, A1 Movers and A2 Flyers, and practical requirements to ensure alignment with A2 Key for Schools and the CEFR. Based on the findings of this research, new speaking assessment scales were proposed.

The proposed new speaking assessment scales met several criteria. These criteria are listed below.

To ensure standardised CEFR alignment:

- descriptors were linked to the planned new CEFR descriptors for young learners where possible.

To ensure alignment with A2 Key for Schools:

- where developmentally appropriate, A1 Movers and A2 Flyers descriptors were aligned with A2 Key for Schools descriptors.

To assist alignment between Pre A1 Starters, A1 Movers and A2 Flyers:

- consistent language and descriptors were used across the assessment scales
- where test design permitted, descriptors showed stepped development, e.g. Band 5 descriptors for A1 Movers roughly aligned with Band 3 descriptors for A2 Flyers.

To ensure examiners have opportunity to be precise:

- a scale of 0 to 5 for each level was used, with half-band scores (e.g. 3.5, 4.5) and descriptors specified for Bands 1, 3 and 5.

To ensure examiners can use the scales consistently:

- descriptors were worded clearly and unambiguously
- descriptors, where possible, used non-technical language.

To promote ambitious use of language by candidates:

- descriptors, where possible, focused on what candidates can do, and avoided negative descriptors that highlighted missing features or errors.

Three criteria were originally proposed as appropriate for the amount and type of language that candidates produce for Pre A1 Starters, A1 Movers and A2 Flyers exams: Vocabulary (& Grammar), Pronunciation, and Interaction. Several elements were also proposed to fall within these key criteria. An overview of the original criteria is shown in Table 1.1.

Key criteria	Elements within criteria
Vocabulary (& Grammar)	Range, control, extent, cohesion
Pronunciation	Individual sounds, (word) stress, intonation
Interaction	Responding, support required, fluency

Table 1.1: Overview of the criteria identified in the original scale development study

In the proposed criteria, Vocabulary (& Grammar) referred to candidates' ability to complete test tasks using the vocabulary and grammatical structures on the Pre A1 Starters, A1 Movers and A2 Flyers syllabus. The ability to use lexis (and structure) is a fundamental component of developing language proficiency, and while exact terminology can differ, 'vocabulary' and 'grammar' are widely recognised terms. As communication at the level of Pre A1 Starters, A1 Movers and A2 Flyers relies primarily on selecting correct lexis and much less on accurate use of structures, Vocabulary was recommended to be presented first. Note that in Table 1.1, Grammar is listed in brackets because it was proposed that for Pre A1 Starters, only 'Vocabulary' is relevant as most utterances are one or two words. For A1 Movers and A2 Flyers, the proposed key criterion was 'Vocabulary and Grammar'.

Regarding the proposed elements within the Vocabulary (& Grammar) criterion, the following elements were proposed:

- range, referring to the degree to which candidates have the vocabulary to deal with the test tasks
- control, referring to the degree to which candidates can (attempt to) use the simple structures which are in the syllabus for the exams
- extent, referring to the length of a candidate's utterances
- cohesion, referring to candidates' ability to join ideas using simple linkers, e.g. 'and' at Pre A1 Starters, 'then' and 'because' at A2 Flyers.

Next, the proposed key criterion 'Pronunciation' referred to candidates' degree of intelligibility, which is of more interest at the level of Pre A1 Starters, A1 Movers and A2 Flyers than accurate use of sounds. The term 'pronunciation' was recommended to be retained because it is widely understood, and retaining its use also maintains consistency with the speaking assessment scales of other Cambridge English Qualifications.

Regarding the proposed elements within the Pronunciation criterion, the following elements were proposed:

- individual sounds, referring to the appropriate use of phonemes and syllables
- (word) stress, referring to the appropriate use of stress within words as well as in sentences, e.g. im-'**por**-tant and not im-por-'**tant**
- intonation, referring to appropriate use of rising and falling pitch.

Finally, the proposed key criterion Interaction referred to candidates' ability to understand or communicate verbal information with another speaker. Note that at the level of Pre A1 Starters, A1 Movers and A2 Flyers this often occurs with the help of pictures and non-verbal communication.

Regarding the proposed elements within the Interaction criterion, the following elements were proposed:

- responding, referring to how the candidate responds to prompts from the examiners, e.g. questions, invitations to speak
- support required, referring to the amount of support the examiner needs to provide so candidates can complete the task
- fluency, referring to the ease, spontaneity and naturalness with which candidates verbally interact with the examiner.

External expert review proposed only minor changes to the proposed assessment scales. Therefore, the next stage of development and validation was started.

DESCRIPTOR SORTING EXERCISE AND INITIAL TRIAL

Twenty-eight experienced teachers of young learners from around the world, who were also experienced Pre A1 Starters, A1 Movers and A2 Flyers examiners, participated in an online sorting exercise. In the exercise, the participants were given a selection of descriptors from the new Pre A1 Starters, A1 Movers and A2 Flyers speaking assessment scales and the current A2 Key for Schools. The participants then identified which level they thought

the descriptors represented: low-Pre A1 Starters, mid-Pre A1 Starters, mid-A1 Movers, mid-A2 Flyers/A2 Key for Schools, high-A2 Flyers/A2 Key for Schools, or all levels. These levels took into account overlap between the exams and represented the full range of achievement (e.g. mid-A1 Movers is the same as high-Pre A1 Starters).

The participants' responses were analysed using many-faceted Rasch measurement (MFRM) and the results informed substantial revisions to the descriptors. Overall, the MFRM indicated that the descriptors covered all levels, and that participants were consistent after taking into account that they had (necessarily) not undergone standardisation training.

Descriptors that were placed into 'wrong' levels or placed inconsistently were typically found to require rewording to improve consistent alignment of descriptors across Pre A1 Starters, A1 Movers and A2 Flyers. These revisions were made and an initial rating trial was conducted to gauge experienced examiners' experiences with the new scales, and to elicit comments from them.

In the initial trial, examiners used the drafts of the new scales to rate existing video of Speaking tests. They reported their scores and completed a questionnaire about their experience using the new scales, and provided written comments and suggestions.

Several themes emerged. Examiners were of mixed opinions about whether half bands were useful at such low levels of language proficiency. Half bands for Pre A1 Starters were consistently reported to be unnecessary, and similar comments were often made for A1 Movers and A2 Flyers. For example:

I was more confident [awarding half bands] for A2 Flyers (and A1 Movers to some extent) but the lack of language produced by candidates at Pre A1 Starters level would seem to fit them into one of the main bands (Band 1, 3 or 5) more immediately.

After review of the descriptors and videos of candidate responses, it was concluded that integer bands of 0 to 5 provided sufficient preciseness without asking examiners to make differentiation at a level of detail not determinable from candidate responses.

Based on examiner requests, wording consistency was further tightened by standardising use of modifying adverbs (e.g. *some, most, all, sometimes, often, almost always*) across the Pre A1 Starters, A1 Movers and A2 Flyers scales.

In addition, several revisions to the Pre A1 Starters speaking assessment scales were made. First, references to grammar at Pre A1 Starters were removed due to observations such as 'grammar at Pre A1 Starters – [there is] not enough language, mistakes are about wrong choice of words or individual sounds/word stress'. Also based on examiners' observations, references to linking words and cohesion were removed from Pre A1 Starters, as Pre A1 Starters candidates only need to link individual words. However, similar references were retained in A1 Movers, as the tasks offer opportunity to begin experimenting with cohesive speech. Finally, examiners noted that most words produced at Pre A1 Starters are monosyllabic, therefore reference to word stress was removed from Bands 1 and 3 of the Pre A1 Starters scale.

Regarding the Interaction key criterion, examiners made useful observations about the need to assess successful reception as well as successful responding at the low levels of language proficiency targeted in Pre A1 Starters, A1 Movers and A2 Flyers, so the element reception was added alongside responding. Examiners also observed that promptness of response was a key aspect of fluent responding at this low level of language proficiency, so the element promptness was added alongside fluency.

TRIALLING FOR PRECISENESS AND CONSISTENCY

The final development activity conducted for the speaking assessment scales was a trial involving 21 experienced Pre A1 Starters, A1 Movers and A2 Flyers examiners located around the world. These examiners used the drafts of the revised speaking assessment scales, and also the pre-revision speaking assessment scales. Sixty videos of candidates participating in Pre A1 Starters, A1 Movers or A2 Flyers Speaking tests were used (20 candidates at each level). The candidates were drawn from East Asia, the Middle East, Europe and Latin America, and their Speaking test videos were selected to represent a range of performances at each level (Pre A1 Starters, A1 Movers and A2 Flyers).

Examiners scored half the total sample (e.g. 30 videos), once with the drafts of the revised scales, and once with the pre-revision scales. To reduce scoring bias arising from using one scale first, a random half of examiners used the pre-revision speaking assessment scales first, while the other half of examiners used the draft revised speaking assessment scales first. A week later, examiners scored the videos again, using the other scales. Table 1.2 shows the linking design for the trial.

Pre A1 Starters	A1 Movers	A2 Flyers	Candidate group	Rater group		
				R1	R2	R3
n = 5	n = 5	n = 5	C1	X	X	X
n = 5	n = 5	n = 5	C2	X		
n = 5	n = 5	n = 5	C3		X	
n = 5	n = 5	n = 5	C4			X
				n = 7	n = 7	n = 7

Table 1.2: Linking design for trial

Using the scores submitted by the examiners, MFRM analyses were conducted to identify scoring consistency among examiners, and also to investigate whether the revised scales were enabling examiners to make more precise diagnoses. Note that the sample size, which was partially determined by logistical factors, was at the lower boundary of acceptability for an MFRM analysis, so results should be interpreted with some caution.

TRIAL FINDINGS

The MFRM analysis results indicated that examiners were using the scales with acceptable consistency, considering they had (necessarily) not yet received standardisation training. Examiner self-perceptions, submitted via an online questionnaire, also indicated satisfaction with the scales, but there were clear requests for standardisation training for the revised scales; this training is mandatory for all existing and new examiners.

The MFRM analysis results also indicated that the revised scales were enabling examiners to differentiate more clearly among stronger candidates. Specifically, fewer candidates were awarded full marks on each of the three key criteria, with marks more spread out over Bands 3, 4 and 5. This change provided the information necessary to conduct alignment studies.

A limitation to note is that, as with operational testing and in line with the high-facility design of the exams, there were very few scores awarded at Bands 1 and 2. Therefore it is not possible to draw strong conclusions about how examiner preciseness has shifted for Bands 1 and 2. However, Figure 1.1 shows that a similarly helpful shift appears to have occurred through use of the revised scales.

Specifically, Figure 1.1 shows that there have been shifts in score assignment, and shows to what extent scores from the revised and pre-revision assessment scales can be expected to overlap. Column 'Measr' shows the measurement scale for each analysis (which was a Rasch scale in 'logits'). Column S.1 represents the pre-revision assessment scale, and column S.2 represents the revised assessment scale. Comparing the band overlap for each score shows, for example, that the low end of pre-revision Pre A1 Starters and A2 Flyers full marks scores (3) now receive '4's on the revised scales rather than full mark scores (5). Likewise, score '1' on the pre-revision A1 Movers assessment scale is split between scores 1 and 2 on the revised assessment scale.

Pre A1 Starters

Measx	S.1	S.2
8 + (3) + (5)		
7 + +		
6 + +		
5 + +		
4 + +		
3 + +		
2 + + 4		
1 + +		
0 * 2 * 3 *		
-1 + +		
-2 + (1) + (1)		

A1 Movers

Measx	S.1	S.2
7 + (3) + (5)		
6 + +		
5 + +		
4 + +		
3 + + 4		
2 + +		
1 + +		
0 * * *		
-1 + +		
-2 + + 1		
-3 + (0) + (0)		

A2 Flyers

Measx	S.1	S.2
10 + (3) + (5)		
9 + +		
8 + +		
7 + +		
6 + +		
5 + +		
4 + +		
3 + +		
2 + + 4		
1 + +		
0 * 2 * 3 *		
-1 + +		
-2 + + 2		
-3 + (1) + (1)		

Figure 1.1: Overall comparability of scores on each Speaking test (all constructs)

Likewise, Figure 1.2 compares candidates' 'fair average' scores, which account for rater severity/leniency and can be considered MFRM 'average' scores for a candidate, for the pre-revision versus the revised scales. The figures show that where on the pre-revision scales there are clusters of responses at the top of the scale, on the revised scales responses are less clustered and more spread out. This is a positive finding, and shows again that the revised scales are allowing examiners to score with more precision.

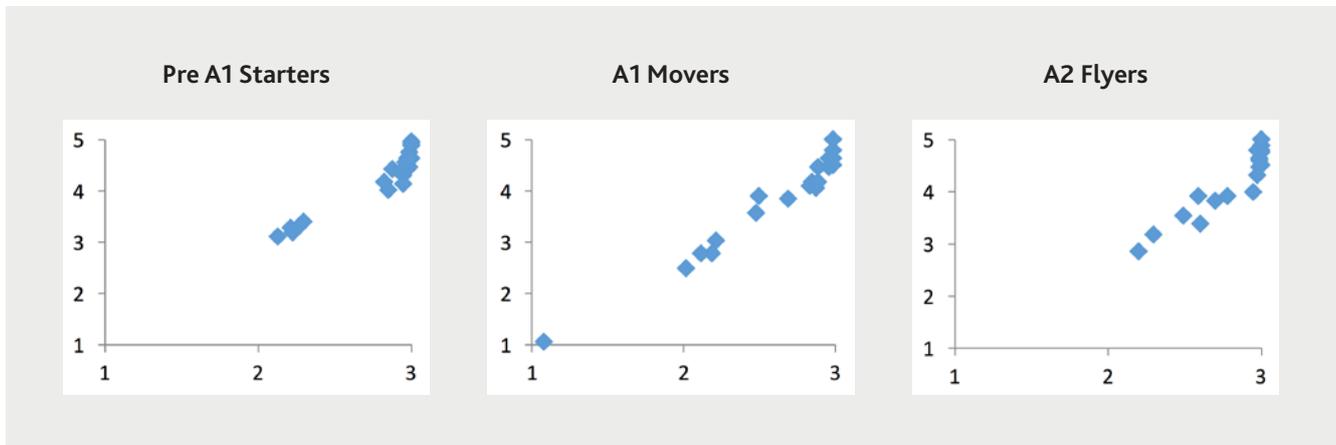


Figure 1.2: Fair average scores for candidates from pre-revision (x axis) and post-revision (y axis) assessment scales

CONCLUSION

In summary, the newly developed speaking assessment scales represent an age-appropriate speaking construct for primary school-aged children who are just beginning to acquire English proficiency. The scales have undergone expert theoretical review and expert practitioner review, and have been empirically tested to ensure the goals of the new scales have been achieved: increased preciseness and continued practical usability.

Chapter 2

Creating an easy-to-understand alignment for Pre A1 Starters, A1 Movers, A2 Flyers, A2 Key for Schools and the CEFR

Mark Elliott, Maggie Dunlop

CONTEXT

The pre-revision Pre A1 Starters, A1 Movers and A2 Flyers exams are reported on a simple scale of one to five shields for each component (Reading/Writing, Listening and Speaking) and there is no overall score. While this system is simple, easy to understand and popular, one of the most frequent queries about Pre A1 Starters, A1 Movers and A2 Flyers before 2018, and a theme from initial research preparing for the revisions, was the relationship between the shield scores at different levels. For example, what is the A1 Movers equivalent of two shields on A2 Flyers? This relationship was not precisely specified in the pre-revision exams; it is simply advised that a candidate scoring a total of 10 or more shields is ready to move to the next level. This is useful information, but a little vague as it does not provide concrete alignment, particularly at the level of individual papers (i.e. Listening, Reading/Writing, Speaking).

In order to provide clearer information about what Pre A1 Starters, A1 Movers and A2 Flyers results mean, one of the main goals of the revisions project was to create an empirically-based mapping between the different levels. In this way, a clear relationship between shield scores on different levels could be established. A further goal was to clarify the relationship between the exams and the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001). The relationship between the pre-revision exams and the CEFR is based on content mapping between the exams and the CEFR. Pre-revision A1 Movers is a test based on content mapping of CEFR A1 and pre-revision A2 Flyers is a test based on content mapping of CEFR A2. No studies were conducted to establish exactly how many shields corresponded to the CEFR A2 boundary.

This approach was useful because it maintained a strong encouragement-oriented ethos for the exams, and minimised any possibility of misusing them (e.g. using results to rank children or to set entry requirements in schools). However, this approach also meant that it was impossible to identify exactly when a candidate had achieved CEFR A2 English proficiency. Over time, internal studies were conducted to establish the nature of the relationship between A2 Flyers and the CEFR, but although the results provided a performance mapping, they did not provide clarity in terms of shields because the CEFR A1/A2 threshold on each paper fell somewhere in the middle of the range of scores covered by a shield, rather than at the shield boundary. For example, for the Reading/Writing paper, the CEFR A2 boundary lay somewhere between three and four shields.

The purpose of the alignment project was therefore to make clear answers to the following questions: How do shields on the Pre A1 Starters, A1 Movers and A2 Flyers exams correspond with each other? How do A2 Flyers shields align with A2 Key for Schools results? Exactly how do the Pre A1 Starters, A1 Movers and A2 Flyers exams align to the CEFR? It should be noted that as a direct result of this alignment process, shield scores on the pre-revision and revised exams are not directly comparable, since aligning the exams naturally involves shifting the boundaries that determine shield scores. In order to carry out this project, three main tasks were identified:

1. Develop new speaking assessment scales that are aligned with each other and the CEFR, and that provide enough precision to identify achievement across CEFR pre-A1, A1 and A2.
2. Ensure task types across Pre A1 Starters, A1 Movers and A2 Flyers are consistent where possible, to improve content alignment between the test.
3. Conduct psychometric scale alignment work to ensure Pre A1 Starters, A1 Movers and A2 Flyers shields are clearly matched to each other, to A2 Key for Schools and the CEFR.

Task 1 was discussed in Chapter 1. Work undertaken for Tasks 2 and 3 is now presented.

INCREASING CONSISTENCY OF TASK TYPES

Some minor changes were made to the Reading/Writing and Listening papers to increase the consistency of task types across Pre A1 Starters, A1 Movers and A2 Flyers. For Listening papers, an opportunity was identified during the review process to improve progression between Pre A1 Starters Part 1, A1 Movers Part 1 and A2 Flyers Part 1, which are all matching tasks that require candidates to draw a line between a picture/word and another picture/word. Therefore, Pre A1 Starters Part 1 now asks candidates to draw lines between names and people on a picture instead of indicating location of objects on a picture. This creates age- and level-appropriate consistency between the three exams, as A1 Movers Part 1 and A2 Flyers Part 1 also ask candidates to draw lines between names and a set of pictures. In addition, A1 Movers Part 3 now asks candidates to match words to a set of pictures instead of drawing lines from days of the week to a set of pictures. This change creates further age- and level-appropriate consistency between A1 Movers and A2 Flyers.

Secondly, the number of questions on the Reading/Writing papers was adjusted to ensure progression between the three exams. Specifically, the purpose was to ensure that for equivalent tasks in each test, lower levels have the same or fewer items as higher levels. For example, three questions were removed from A1 Movers Part 5 to ensure the task has the same number of items as the equivalent task in A2 Flyers (i.e. post-revision Part 5). In addition, one question was removed from A1 Movers Part 1, and one question was removed from A1 Movers Part 3 (pre-revision Part 4).

	Pre-revision	Post-revision
Pre A1 Starters	25 (25)	25 (25)
A1 Movers	40 (40)	35 (39)
A2 Flyers	50 (50)	44 (48)

Table 2.1 Number of questions and marks* on Reading/Writing papers

* Marks in brackets. 'Marks' includes the extended writing tasks that receive multiple marks for one question.

SCALE ALIGNMENT TASK

The larger piece of work in the alignment project was the scale alignment. The goal of the scale alignment task was to align the number of shields across Pre A1 Starters, A1 Movers and A2 Flyers, and to provide a consistent, transparent alignment of shield numbers to the CEFR. A broad alignment of shield scores to the Cambridge English Scale was also determined, since the Cambridge English Scale is essentially an enumeration of the CEFR, but this was not a primary goal, and alignment is not specified beyond the range of marks covered by a CEFR level. For example, a score of three shields on A2 Flyers corresponds to five shields on A1 Movers, CEFR Level A1 and a Cambridge English Scale score in the range of 100–119. Full details of the alignment design are shown in Figure 2.1 below.

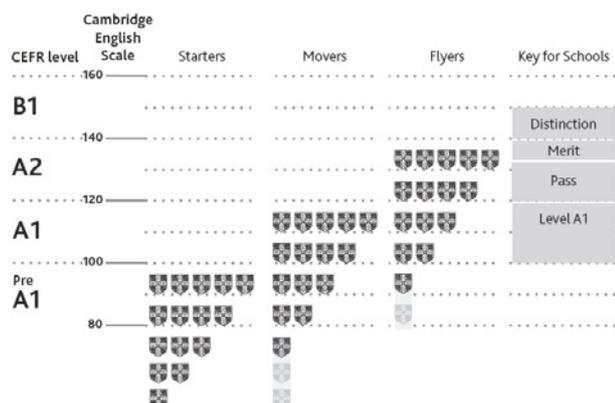


Figure 2.1: Design of alignment between Pre A1 Starters, A1 Movers and A2 Flyers, the CEFR and the Cambridge English Scale

The boundaries between different CEFR levels in Cambridge English Qualifications (A2 Key/A2 Key for Schools, B1 Preliminary/B1 Preliminary for Schools, B2 First/B2 First for Schools, C1 Advanced and C2 Proficiency) are fixed points on the measurement scale used by Cambridge Assessment English across all three exams. This scale is based on a psychometric modelling method for creating measurement scales from test scores known as the Rasch model (Rasch 1960/1980).

In order to link the Listening and Reading/Writing shields to the CEFR and to each other, researchers:

- determined the CEFR A1 and A2 boundaries for A2 Flyers Listening and Reading/Writing by conducting a study comparing data from candidates taking A2 Flyers and A2 Key for Schools
- created a single measurement scale for Pre A1 Starters, A1 Movers and A2 Flyers so that the boundaries from A2 Flyers could be applied to A1 Movers and Pre A1 Starters
- developed a system for linking versions over time so that the initial boundaries will remain stable.

Determining CEFR A1 and A2 boundaries

The first task was to determine the CEFR A1 and A2 boundaries for A2 Flyers on the Common Cambridge Scale. For this task, the results of a previous alignment study between A2 Flyers and A2 Key for Schools were used. In that study, live A2 Flyers candidates were invited to take A2 Key for Schools, and live A2 Key for Schools candidates were invited to take A2 Flyers to compare performance on the two exams. This design is referred to as *counterbalanced* – the live A2 Flyers candidates balance the live A2 Key for Schools candidates, since differences in motivation and preparedness between the live test and the additional test should cancel out between the two groups. All candidates were entered as live candidates for both exams and received official results and certificates. The candidates' responses across the two exams formed the basis of the comparison; data from other live candidates who did not take part in the study boosted the precision of the analysis.

Because the A2 Flyers boundaries were to be linked to the CEFR via A2 Flyers' relationship with A2 Key for Schools, it was first important to clarify that the constructs of A2 Flyers and A2 Key for Schools are sufficiently close – in other words, that both essentially test the same concepts of Listening and Reading/Writing. This was investigated statistically by two methods: plotting raw test scores from A2 Flyers and A2 Key for Schools against each other, and from measures of item fit to the Rasch model:

1. Raw scores for the Reading/Writing components on each test were plotted against each other and a linear regression model (a line of best fit) was constructed. This gave an R² (a measure of correlation) of 0.72 – values closer to 1 indicate a better fit – which represents a strong correlation. The correlational evidence was less clear for the Listening components as the R² for Listening raw scores was notably lower (R² = 0.41, which is moderate), possibly due to differences in response format between A2 Flyers and A2 Key for Schools.
2. All items for both Listening and Reading/Writing were included together in a Rasch model and Rasch statistics describing the fit of the data to the theoretical model were produced. These statistics are known as 'infit mean square values' and should be close to a value of 1 – higher values are particularly problematic since they indicate unexplained 'noise', which typically means that an item is testing something different from the other items. In this study, the fit statistics for all items were between 0.7 and 1.3, which indicates that the items on both tests fitted a single model acceptably.

These results indicate that we can be confident the two tests are not testing significantly different constructs.

After pooling the items to analyse them together, the A2 Flyers items were calibrated on the A2 Key for Schools scale through the combined Rasch analysis. With the A2 Flyers items on the A2 Key for Schools scale, the A2 Key for Schools boundaries for CEFR Levels A1 and A2 could also be applied to A2 Flyers Listening and Reading/Writing. In this way, A2 Flyers Reading/Writing and Listening were mapped to the CEFR. Shield numbers were then determined according to the relationship described in Figure 2.1 – by dividing each CEFR level into two and, for example, awarding four shields on A2 Flyers for the bottom half of A2 and five for the top half of A2.

Putting Pre A1 Starters, A1 Movers and A2 Flyers Listening and Reading/Writing on a single scale

The next task was to ensure that the boundaries for Pre A1 Starters and A1 Movers were consistent with the boundaries for A2 Flyers. This was done by an analysis of the functioning of items from across the three levels together, similarly to the way A2 Flyers and A2 Key for Schools were aligned. In this case, the same A2 Flyers version used in the A2 Flyers/A2 Key for Schools alignment study and the corresponding Pre A1 Starters and A1 Movers versions were used to create a number of 'hybrid' test versions. They were hybrids because they were made from different combinations of tasks from tests at adjacent levels:

- three Pre A1 Starters/A1 Movers Reading/Writing hybrids
- three Pre A1 Starters/A1 Movers Listening hybrids
- three A1 Movers/A2 Flyers Reading/Writing hybrids
- three A1 Movers/A2 Flyers Listening hybrids.

These hybrid versions were administered to cohorts of around 300 candidates each, then combined into one data set for analysis together with regular sets of responses from all standard live administrations of the paper, totalling approximately 2,000 candidates per component. By determining the relative difficulty of the A1 Movers items compared to the A2 Flyers items, the A1 Movers items – and thus the A1 Movers scale – were aligned to the A2 Flyers scale, and similarly for Pre A1 Starters.

Once the scales for Pre A1 Starters, A1 Movers and A2 Flyers were aligned, the same CEFR A1 and A2 boundaries drawn from A2 Key for Schools onto A2 Flyers were applied to the Pre A1 Starters and A1 Movers scales, meaning that a consistent standard could be applied across all three levels.

Maintaining alignment

Finally, in order for the alignment which had been established to produce any long-term benefits, a robust means of maintaining alignment across multiple test versions was necessary. Therefore, in the revised Pre A1 Starters, A1 Movers and A2 Flyers exams, each new test version contains 'anchor tasks', which are tasks which have already appeared in a previous live test. These tasks have already been analysed and so the task items have known difficulty values, and these have been previously placed on the same scale as the CEFR boundaries. Because the item difficulty values are known and are on the same scale as the CEFR boundaries, using anchor tasks means that new test tasks can also be placed on the same scale. In this way, standards can be maintained over time by creating an unbroken chain of anchor tasks from the first live versions.

SPEAKING

The alignment study also compared speaking scores obtained on A2 Flyers with speaking scores obtained on A2 Key for Schools; scores on the two tests were plotted against each other as for the other components.

The study found that it was not possible to generate a robust boundary for A2 performance on the A2 Flyers Speaking test. This was because of a 'ceiling effect' on A2 Flyers: some candidates who were awarded maximum marks on the pre-revision A2 Flyers Speaking test did not achieve A2 level on the A2 Key for Schools Speaking test. This issue was addressed by developing new speaking assessment scales (see Chapter 1). The new scales facilitate greater precision of marking speaking performance, thereby providing the information necessary to calculate CEFR A1 and A2 level performance.

CONCLUSION

For the Listening and Reading/Writing components, a series of statistical studies were carried out to construct linked scales across Pre A1 Starters, A1 Movers and A2 Flyers and to determine CEFR boundaries. These studies allowed the creation of a clear, aligned reporting system with a precise relationship between shield scores across levels and to the CEFR. The test design process ensures that this linking can be maintained over time. For Speaking, the revised scales ensure that CEFR linkage is built in and that precision is increased at the top end of the score range.

Chapter 3

New writing tasks for young learners: A1 Movers and A2 Flyers

Kathryn Davies, Maggie Dunlop

INTRODUCTION

In the pre-revision A1 Movers and A2 Flyers exams, candidates were asked to do only a limited amount of writing; just one to three words for A1 Movers and one to four words for A2 Flyers. Moreover, all writing consisted of either selecting text from cloze options or from a reading text. Although the reading comprehension demonstrated on texts used in the exams provided insight into candidates' understanding of writing conventions, there was no specific opportunity for candidates to produce any extended, independent writing. This limited the extent to which the tests could draw meaningful inferences about candidates' English writing ability.

It therefore became a priority during the revision project to develop writing tasks that would enable A1 Movers and A2 Flyers candidates to be properly rewarded for demonstrating their emerging English writing skills. This was particularly an opportunity to better support classroom teachers in their communicative teaching practices. An opportunity was identified when the pre-revision A1 Movers and A2 Flyers Part 2 was identified for replacement (discussed in Chapter 4), and a space on the papers opened up. The development studies addressed three main questions: What writing should be expected at these low levels of English proficiency and young ages? What tasks will elicit appropriate written language? How should writing be assessed?

WHAT TO EXPECT

Pre A1 Starters already included a productive writing task (Part 5). This task was retained unchanged in the post-revision tests. As appropriate for Pre A1 Starters level – pre-A1 on the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) – candidates are required to demonstrate their writing skills at single-word level only for this task. The revision project therefore sought to introduce productive writing at A1 Movers and A2 Flyers levels to enable the measurement of extended writing skills.

According to the CEFR (Council of Europe 2001), A1 level language learners (the upper achievement category of A1 Movers) can write simple sentences. At A2 level (the upper achievement category of A2 Flyers), language learners can write simple connected texts. Relevant CEFR descriptors are shown in Figure 4.1.

Descriptors for writing at A1 level

- Can use simple words and phrases to describe certain everyday objects (for example the colour of a car, whether it is big or small)
- Can write simple isolated phrases and sentences

Descriptors for writing at A2 level (A2 Flyers)

- Can link groups of words with simple connectors like 'and', 'but' and 'because'
- Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points

Figure 4.1: Relevant CEFR descriptors for writing

A1 MOVERS TASK DEVELOPMENT: HOW TO ELICIT CEFR PRE-A1/A1 LEVEL WRITING

A new writing task was developed for A1 Movers, four versions of which were trialled across a total of 891 A1 Movers-level candidates (approximately 200 learners per version, each cohort representative of the target candidature in terms of geographical spread, age and CEFR level). The four different versions were created to determine which task elicited the most amount of, and most appropriate, language.

Candidates were prompted to write three to four sentences giving their name, age, likes and dislikes, in line with expectations of CEFR A1 level young learner candidates' writing skills (see Table 4.1). The intention of the new task was to encourage candidates to produce extended writing and to provide an opportunity for personalisation.

Version	Model	Prompt
A	Jack likes ... but he doesn't like ... Lilly likes ... but she doesn't like ...	My name is ... I like ... I like ...
B	Jack likes ... but he doesn't like ... Jack likes ... but he doesn't like ...	My name is ... I like ... I like ...
C	Jack is 9 years old. He likes ... His favourite colour is ...	My name is ... I'm ... years old. I like ... My favourite colour is ...
D	Sally is 9 years old. She likes ... She doesn't like ...!	My name is ... I'm ... years old. I like ... I don't like ...

Table 4.1 Summary of writing prompts trialled

Results showed that the model text employed in versions A and B (see Table 4.1) – vital scaffolding in order to explain the requirements of the task – appeared to be setting candidates up for failure. As exemplified below in Figure 4.2, some candidates were directly lifting the (third person singular) verb form from the model and using it incorrectly in the first person. There was no way of knowing whether, if set up differently, candidates would have been able to produce the same meaning more accurately or not.

Now complete three sentences about you.

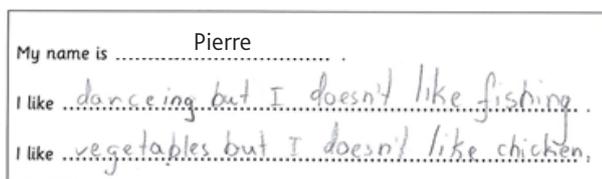


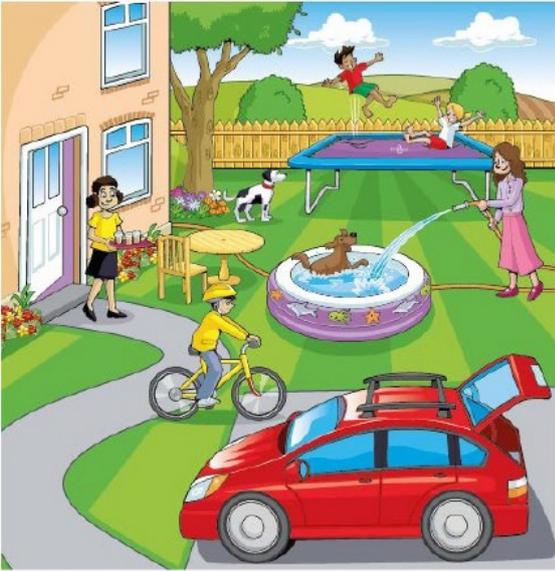
Figure 4.2: Example response that copies third person example into a first person answer

Trial results revealed a second issue in that some candidates were not expanding their answers and were therefore not demonstrating the full extent of their writing skills. In addition, experts and experienced practitioners participating in task development raised concerns about candidates having to write their name and age. It was noted that writing one's name and age does not provide evidence of emerging ability to write independent, extended text. Finally, the appropriate criterion for marking 'age' responses was unclear. For example, would candidates gain more marks for writing their age in letters (e.g. 'ten') and fewer if it was written numerically (e.g. '10')? This would not be a fair criterion, as '10' is a correct response in the context, but '10' is also not a written word.

In response to the issues identified, a new writing task was developed, trialled and approved. The task still consists of several questions (rather than demanding a cohesive piece of text as in A2 Flyers), and still provides a lot of support to candidates. However, in the new writing task, candidates' ability to produce writing ranging from words and phrases to grammatically complete sentences about a picture is tested. The task is composed of six questions and is designed to provide greater freedom and gradually reduce scaffolding as candidates progress through the task. Candidates complete sentences (Questions 1 and 2), respond to direct questions (Questions 3 and 4), and write sentences (Questions 5 and 6) about a picture. Crucially, the free-response items (Questions 5 and 6) give candidates the opportunity to really demonstrate their sentence-writing skills, thus allowing fuller assessment of the writing construct, and potentially create positive washback effects. Moreover, the task as a whole elicits far more language than the original A1 Movers task. An example of the new task is shown in Figure 4.3. An example set of responses is shown in Figure 4.4.

Part 6
- 6 questions -

Look and read and write.



Examples

The dog is playing in some water

What's the mother wearing? a purple shirt

Questions

Complete the sentences.

1 The boy on the bike is wearing gray

2 One dog is brown and one dog is

Answer the questions.

3 What's the girl in the black skirt doing?
.....

4 Where are the flowers?
.....

Now write two sentences about the picture.

5

6

Figure 4.3: Example of new A1 Movers productive writing task

Part 6
- 6 questions -

Look and read and write.

Examples

The dog is playing in some water

What's the mother wearing? a purple shirt

Questions

Complete the sentences.

1 The boy on the bike is wearing grey

2 One dog is brown and one dog is white and black

Answer the questions.

3 What's the girl in the black skirt doing?
water

4 Where are the flowers?
next to the tree

Now write two sentences about the picture.

5 The car is red

6 The boy that is jumping has red shirt

Figure 4.4: Example response to new A1 Movers productive writing task

A1 MOVERS ASSESSMENT SCALE DEVELOPMENT: HOW TO ASSESS CEFR PRE-A1/A1 LEVEL WRITING

Initially a holistic assessment scale using a single rating scale was proposed for the new A1 Movers picture description task, despite the task consisting of six discrete items. This was developed after the work on the A2 Flyers initial assessment scale (see below) and drew on what was learned from A2 Flyers. (The A1 Movers assessment scale is presented here first for logical consistency.) The initial A1 Movers assessment scale was developed using sample responses to the new task. The samples were analysed and categorised into four groups according to their relative strength as a piece of writing. The boundary points between each group were identified and defined, and then used as the basis for formulating a rating scale (0 to 5). The scale was then revised following consultation with experts, piloting and subject to further minor revisions before being trialled on a larger scale. In total 75 A1 Movers responses were collected from schools in Latin America, East Asia and Europe, and these responses were scored by 41 markers based around the world.

Results of a many-faceted Rasch measurement (MFRM) analysis indicated that the A1 Movers assessment scale differentiated between stronger and weaker candidates; there was a spread across score bands. However, the markers found it challenging to use a holistic mark performance over discrete items. Correlation analyses were performed to estimate the strength of agreement between markers' scores and the results indicated that, linked to the challenges markers faced in interpreting it, scorer consistency was insufficient.

As a result, a second assessment scale requiring markers to score items discretely was developed. Questions 1 and 2 (sentence completion) were to be marked 'correct' and receive a maximum of 1 mark each where the response was deemed to provide an accurate representation of the picture. For each of Questions 3–6, candidates could score a maximum of 2 marks: 1 mark for a response which was comprehensible; and 1 mark for a response which was an accurate representation of the picture. Accuracy in grammar and spelling was not to be marked; comprehensibility was sufficiently accurate use of language at this language proficiency level.

A second trial was conducted to determine whether this revised assessment scale improved marker consistency. The trial used the same set of responses as the first trial. Each of these responses was scored by 36 markers.

The results of an MFRM analysis indicated that Questions 3–6 differentiate between candidates. Item difficulty was similar for all four questions, which came out at an appropriate level of difficulty for an A1 test. Scorer consistency was good. For example, for Questions 1 and 2, on average 87% of markers awarded the same score. For Questions 3–6, on average 83% of markers awarded the most frequently given score (mode) with a further 15% awarding a score within one mark of the mode. Some difference of interpretation is to be expected among human markers (McNamara 1996).

A2 FLYERS TASK DEVELOPMENT: HOW TO ELICIT CEFR A1/A2 LEVEL WRITING

For the new A2 Flyers task, two different types of writing tasks were trialled with A2 Flyers-level candidates: all four versions of the original information task designed for A1 Movers, which allowed candidates to personalise their responses, and two versions of a new story task (see Figure 4.5). The story task was designed to test candidates' ability to produce a short discourse-level written text by eliciting a connected text based on a series of three pictures that depicted a simple story. The images for one version depicted events during a game of beach volleyball, while the other depicted events during an astronaut's visit to an alien planet.

A total of 814 A2 Flyers-level candidates trialled both task types. Approximately 200 candidates completed each of the four A1 Movers information task versions (each cohort representative of the target candidature in terms of geographical spread, age and CEFR level). In addition, approximately half of candidates also completed the beach volleyball story, and a half of candidates completed the astronaut story.

Candidates' responses on the (A1 Movers) information task were compared with those produced by the A1 Movers trial candidates. As A2 Flyers candidates generally successfully completed all the prompts, it was clear that the information task did not differentiate between stronger and weaker A2 Flyers-level candidates. This task was therefore confirmed as providing insufficient scope for stronger A2 Flyers candidates to demonstrate their ability. In contrast, despite offering less scope for personalisation, the story task clearly emerged as the more appropriate task for the level. Upon reflection, this is as expected given that the task aligns more closely with the relevant CEFR A2 level Can Do statements. The story-writing task was therefore selected as the new productive writing task for A2 Flyers.

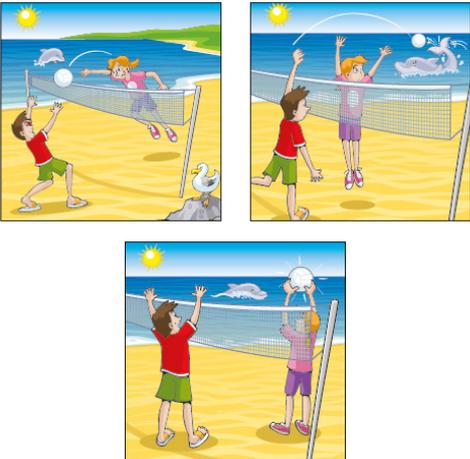
<p>Part 7</p> <p>Look at the three pictures. Write about this story. Write 20–30 words.</p>  <p>One day Tom and Kate went to beach played in volleyball. Tom throw strongly ball and Kate didn't catch. But suddenly dolphin hit ball. In the end Kate caught ball.</p>	<p>Part 7</p> <p>Look at the three pictures. Write about this story. Write 20–30 words.</p>  <p>The astronaut find Aliens. the aliens and the astronaut start talking. The one alien invited to eat a astronaut because the astronaut have to eat and after the astronaut and the alien go to their houses.</p>
---	--

Figure 4.5: Examples of new A2 Flyers productive writing task, with example responses

A2 FLYERS ASSESSMENT SCALE DEVELOPMENT: HOW TO ASSESS CEFR A1/A2 LEVEL WRITING

In order to create an assessment scale that fairly and consistently describes children's responses to the new A2 Flyers task at CEFR A1/A2 levels, 10 experienced markers were provided with 40 responses to the new story task and asked to identify features of lower and higher achieving writing following an empirical, binary-choice boundary definitions (EBB) scales approach (Fulcher, Davidson and Kemp 2011, Upshur and Turner 1995, 1999). This required markers to sort appropriate and authentic writing samples into 'better' and 'poorer' groups and then flesh out the criteria that led them to their decisions.

An initial draft assessment scale was developed. It consisted of a 6-point scale (0–5) with a 0 score available where candidates left a blank page. Task completion, i.e. addressing all three pictures, emerged as a determining factor for distinguishing between stronger and weaker responses at the broadest level. To differentiate between stronger responses, linking words and explicit cohesive language emerged as a key indicator. Among weaker responses, relevance of vocabulary and the ability to produce comprehensible phrases and/or sentences were key differentiators.

Assessment experts then compared the draft assessment scale with the constructs to be assessed, to ensure theoretical soundness and to ensure clear alignment with the CEFR. This activity was an adaptation of the EBB methodology, designed to ensure achievement-based assessment criteria were established rather than purely comparative criteria, which would otherwise have been a limitation of adopting a strict EBB approach. The draft assessment scale was then trialled among the same 10 markers for a second time and further refined, mostly to clarify how criteria were expressed rather than adjusting the criteria.

The assessment scale was then trialled. In total, 41 markers located worldwide scored 100 responses collected from schools in East Asia, South Asia, Latin America and Europe. The responses were selected to represent a range of responses distributed across the English achievement continuum and across the four regions.

Trial results indicated the assessment scale differentiated between stronger and weaker responses. However, the assessment scale also needed to further differentiate performance more consistently at the upper end of the scale.

The assessment scale was therefore revised accordingly and further large-scale trialling involving 36 markers was carried out. The second trial used the same set of responses ($n = 100$) as the first trial, but 13 responses were removed to be used as exemplars, thereby reducing the number of trial responses to 87.

The results of an MFRM analysis indicated that the revised A2 Flyers assessment scale differentiated among candidates with responses spread across Bands 2–5. There were many more scores of 5, because some responses exceeded CEFR A2-level expectations for writing. The assessment scale also performed acceptably in terms of scorer consistency. Priming markers to focus on meaning rather than grammatical or spelling accuracy was found to be key to setting the standard at the appropriate level.

CONCLUSION

As a result of trialling, the new writing tasks enable Pre A1 Starters, A1 Movers and A2 Flyers exams to more fully assess the construct of children's emerging English as an additional language writing skills. It is hoped that including these tasks will equip teachers to feel confident in teaching a full range of practical English skills.

For the A1 Movers writing task, trialling showed that the greater freedom and gradual reduction in scaffolding offered as candidates progress through the task gives them the opportunity to demonstrate their writing skills, while supporting weaker candidates. The accompanying assessment scale was found to be reliable and easy to understand. Therefore, the new task was added to the paper as Part 6, replacing the pre-revision yes/no task (Part 2).

Trialling of the A2 Flyers story-writing task showed that the task is clearly achievable for A2 Flyers candidates at CEFR A1/A2 levels of English proficiency. The story task was observed to align closely with the relevant CEFR Can Do statements, which states that A2 level users can link simple sentences in order to tell a story. Results also showed that the assessment scale consistently facilitated differentiating between responses demonstrating CEFR A1 and A2 levels of English writing proficiency. Therefore, the new task was added to the paper as Part 7, replacing the pre-revision yes/no task (Part 2).

Chapter 4

Revising tasks to ensure children have the most opportunity to demonstrate their English skills

Sarah Albrecht, Kathryn Davies, Chris Cullen

INTRODUCTION

As part of the revisions, reviews were conducted to identify potential changes to Pre A1 Starters, A1 Movers and A2 Flyers Listening, Reading/Writing and Speaking tasks. The question asked was: How could tasks be adjusted to ensure children have the best chance possible to demonstrate their English skills?

Answering this question needed to take into consideration the target candidature's level of cognitive development (Field 2011, 2013). Field notes that in pre-adolescence, a child faces the following limitations which will impact on their acquisition and use of a second language, specifically their ability to perform strategically in a test:

- primary-aged children are often not able to match the most effective reading or listening techniques to a given text type or task
- they struggle to switch attention between different sources of information and apply that information appropriately
- their reasoning is such that they might not always recognise inconsistency, and they can be intolerant of ambiguity, both of which may in turn affect their problem-solving skills
- they also find it difficult to distinguish between real and imagined situations, which combined with limited world knowledge impedes their ability to make inferences as listeners and readers
- additionally, their memories are not yet fully developed, which restricts their ability to hold language in their minds for any length of time.

Due to these cognitive characteristics, proposed changes were specifically designed to replicate as far as possible candidates' use of language in the real world, and to avoid cognitive reasoning which may distract candidates from showing their English skills.

Proposed changes consisted of minor revisions to tasks, and the removal of selected tasks (which also made room for the new writing tasks described in Chapter 3). This chapter helps readers understand how these changes were principled and designed with Pre A1 Starters, A1 Movers and A2 Flyers candidates in mind. Note that the changes described here are in addition to the adjustments to tasks described in Chapter 2, which improved task alignment between Pre A1 Starters, A1 Movers and A2 Flyers.

LISTENING TASKS

A review of the Listening test tasks indicated very few changes that needed to be made, as the tasks elicited appropriate use of language, and were popular with candidates. Only one minor change was made, to A1 Movers and A2 Flyers Listening Part 5. Pre-revision, these tasks required candidates to listen to a conversation between two people and follow the instructions. For each question, candidates would be required to either colour something, draw something, or write a word according to the instructions given. On the revised paper, these features remain mostly unchanged. However now, candidates are no longer required to draw. Instead, they will be asked to write an additional word. This change was made because drawing ability is not assessed in the tests, and asking children to draw took their time and attention away from using English in this relatively short assessment period.

READING/WRITING TASKS

A review of the pre-revision Reading/Writing test tasks indicated very few changes that needed to be made, as the tasks generally elicited appropriate use of language, and were popular with candidates. However, four minor changes were identified that made it easier for children to show their true English skills. Changes to Pre A1 Starters tasks are discussed first, followed by changes to A1 Movers tasks, and finally changes to A2 Flyers tasks are discussed.

A minor change was made to grammar used in Pre A1 Starters Reading/Writing Part 1 so the task better reflected typical English language use, and also offered broader coverage of basic English syntax. On the revised task, sentences can now include plural items (for nouns that often occur as plurals), as well as single items. Specifically, on the pre-revision paper, all sentences would start 'This is a/an ...'. Now, questions can start 'This is a/an ...' or 'These are ...' if the item often occurs in groups or plural form in real life, e.g. 'These are trousers.' 'These are grapes.' This change ensures that words are used in natural ways. Examples of words which can now be included in the task are *glasses, grapes, meatballs, shorts* and *trousers*.

On Pre A1 Starters Reading/Writing Part 4, the genre of the text was altered to a genre more familiar to primary school-aged children. On this task, candidates complete a text using some words provided. Each word provided is accompanied by a picture, to help candidates understand the word. On the revised paper, these features remain unchanged. However, on the pre-revision paper, the text genre used for this task was a riddle. Expert review noted that the riddle genre is not commonly taught to primary school-aged children, and moreover the 'problem' presented did not require solving (the answer was always provided). It was noted that these text features could impact candidates' ability to understand the text and therefore complete the task. As a result, text genre was changed to an age-appropriate factual text.

Pre-revision A1 Movers and A2 Flyers Reading/Writing Part 2 were identified for possible replacement. These tasks required candidates to look at a picture depicting a scene (e.g. children playing in a park), and then read several sentences and write 'yes' if the sentence was true to the picture, or 'no' if the sentence did not describe the picture. Psychometric analyses indicated that items on the task did not consistently differentiate between stronger and weaker candidates. It was hypothesised this was due to the binary response options and task design facilitating guessing. As new writing tasks were subsequently identified as an additional priority, these tasks were removed from the papers.

Finally, for A2 Flyers Reading/Writing Part 3 (pre-revision Part 4) the order in which information is presented has been modified. On the pre-revision task, the story was presented first, followed by the answer options. Now, the answer options appear first, followed by the story. The last question ('Choose the best title for the story') still appears after the story. This change was made because expert reviewers identified that this format should be cognitively less demanding for candidates.

SPEAKING TASKS

A review of the Speaking test tasks indicated very few changes were necessary, as the tasks elicited appropriate amounts and type of language, and are popular with candidates and educators. Several minor changes were considered, including eliciting candidates' name and age, labelling storyboards (which are a series of pictures that depict a simple story in A1 Movers and A2 Flyers Speaking tests) and reducing the number of images in storyboards, merging Pre A1 Starters Parts 1 and 2, and using a 'tell me' question, e.g. 'Tell me about this box.' (Speaking examiner points to a box on the scene sheet). These changes were trialled with 59 Pre A1 Starters, 60 A1 Movers and 45 A2 Flyers candidates in several countries. Feedback was provided by participating examiners regarding candidates' comments, plus examiners' own observations and comments.

Eliciting candidates' name and age was well received by examiners and candidates, and response rates indicated the information elicited at each level was appropriate: name for Pre A1 Starters candidates, name and age for A1 Movers, and name, surname and age for A2 Flyers. This proposed change was accepted.

Labelling storyboards in the A1 Movers and A2 Flyers tests was likewise well received by examiners and candidates, and this proposed change was accepted. However, analysis of trial Speaking tests and examiner feedback showed that reducing the number of images in the storyboard reduced the amount and quality of language produced, and resulted in candidates requiring additional supportive questions. In addition, some children had more difficulty understanding the story in the reduced format. This proposed change was therefore not implemented.

Merging Pre A1 Starters Parts 1 and 2 was originally proposed to reduce the number of purely listening activities on the Speaking test. Trialling indicated that examiners remained confident they had enough information to award marks, and candidates were still often able to successfully complete test tasks. This proposed change was therefore accepted.

Finally, the 'tell me' question was found to provide additional opportunities for candidates to produce speech, and was included for Pre A1 Starters, A1 Movers and A2 Flyers. This question was new for Pre A1 Starters. Pre A1 Starters candidates benefit from concrete prompts, so the question was placed in the Pre A1 Starters task where candidates respond to questions about a scene (post-revision Part 2), so they have a picture to help them form a response. For A1 Movers and A2 Flyers, the question already existed in a conversation task at the end of the test (Part 4), so they could use the opportunity to talk about topics familiar to themselves. The review process indicated that candidates at A1 Movers and A2 Flyers levels were able to handle the question in this prompt-free and personalised context, so its location was retained.

Chapter 5

Updating the Pre A1 Starters, A1 Movers and A2 Flyers vocabulary lists

Lynne Stevenson

INTRODUCTION

This chapter describes the process by which the vocabulary lists for the Pre A1 Starters, A1 Movers and A2 Flyers exams were updated. The vocabulary lists are important guides for both teachers of young learners and Pre A1 Starters, A1 Movers and A2 Flyers exam paper writers regarding the English vocabulary which young learners need to move through the Common European Framework of Reference for Languages (CEFR, Council of Europe 2001) Levels pre-A1 to A2. The lists are publicly available, appearing in the *Pre A1 Starters, A1 Movers and A2 Flyers Handbook for Teachers*.

An update to the vocabulary lists was undertaken as part of the Pre A1 Starters, A1 Movers and A2 Flyers revisions project because English language vocabulary evolves constantly and organically. To ensure the exams are assessing children's ability to use current English language, the vocabulary lists need to contain up-to-date vocabulary.

Research carried out in 2014 found that the majority (about 75%) of responding teachers, administrators and parents felt the size of the vocabulary lists to be 'just right' for Pre A1 Starters, A1 Movers and A2 Flyers. The revisions project therefore only considered essential additions that would be required to keep the wordlists reflecting current usage, and in line with vocabulary breadth and depth expectations at CEFR pre-A1 to A2 levels.

REVIEW PROCESS

The pre-revision vocabulary lists consisted of 466 words for Pre A1 Starters, 824 words for A1 Movers and 1,295 words for A2 Flyers. Note that the word count for A1 Movers included Pre A1 Starters words, and the word count for A2 Flyers included Pre A1 Starters and A1 Movers words. The vocabulary list updating process was based on the following question: 'What vocabulary should be added and removed?' Selections were informed by the kinds of English vocabulary children were observed using, identification of vocabulary that was becoming obsolete, and identification of vocabulary with sufficient orthographic transparency (i.e. the spoken sounds clearly resemble the written letters).

Changes suggested by Pre A1 Starters, A1 Movers and A2 Flyers exam paper writers are routinely recorded. To take a first look at the question 'What vocabulary should be added and removed?' an initial review of the vocabulary lists considered suggestions made since the last vocabulary list review in 2011, as well as recommendations from experienced, practising teachers of young learners. The initial review identified 32 new words to be added to the Pre A1 Starters vocabulary list, 119 new words to A1 Movers and 190 new words to A2 Flyers. The proposed new words for A2 Flyers were divided into productive ($n = 47$) and receptive ($n = 143$) vocabulary. 'Productive' meant words candidates would be expected to speak (and/or write), while 'receptive' meant words candidates would only be expected to understand.

The next step was to assess the extent to which:

1. Children produce the proposed additions.
2. Children produce words or structures not in the pre-revision vocabulary lists or proposed additions.
3. Separate vocabulary lists for receptive and productive vocabulary are feasible at A2 Flyers level.

To do this, a vocabulary bank was created using 256 transcribed Speaking tests, 66 Reading/Writing tests and 66 Listening tests that covered all three levels equally and had been in use from 2009 to 2014. The vocabulary bank included both listening and reading texts from test papers, speaking examiner speech, and candidate written and spoken responses.

The findings showed that:

1. Children at Pre A1 Starters level produced 16 out of the 32 proposed words; children at A1 Movers level produced 61 out of 119 proposed words; children at A2 Flyers level produced 14 out of 47 words from the proposed productive list and 37 out of 143 from the proposed receptive list. These findings informed the next stage of analysis (see below), in which proposed changes were further evaluated.
2. Children at all levels produced words in speaking which were not on the vocabulary lists of their test. Examples included *accidentally, beauty, comfortable, except, mess, owner, rush, shiny, valley*. This finding showed that the Speaking tests elicit personalised and creative responses, and facilitate children to access their own English repertoire to communicate ideas and show their English skills.
3. Although the report from this research recommended that the A2 Flyers vocabulary list should be separated into productive and receptive vocabulary, more research was required to determine how teachers would use this separation. There was a possibility that the receptive/productive division would have encouraged a dichotomy, which could lead to teachers feeling compelled to teach the different categories in artificially different ways. Therefore, the revised A2 Flyers vocabulary list, like the A1 Movers and Pre A1 Starters, does not differentiate between productive and receptive vocabulary.

A list of proposed additions and moves from one test level to another was created, and following this stage experienced test paper writers took part in a workshop to finalise the additions and movement of words.

Criteria followed included:

- Orthographic depth and transparency (i.e. the degree to which the most basic elements of written language (graphemes) reflect the most basic elements of spoken language (phonemes) and therefore the facility of predicting the pronunciation of words from the way they are spelled. English is considered orthographically deep and the need to not overload young learners with too much lexical processing was taken into consideration.)
- Enhancing existing lexical sets with the inclusion of words considered to be lacking, for example, adding *bear* and *zebra* to an existing set of animals.

For a complete list of words which were added, moved or removed see the Appendix. Table 6.1 and Table 6.2 below provide a summary of these changes.

	Pre A1 Starters	A1 Movers	A2 Flyers
New words added	36	41	88
New names added	5	3	3
Words received from other levels	7	16	1
Words given to other levels	0	6	18
Words removed completely from combined vocabulary lists	6	7	26
Total increase of words	42	47	48

Table 6.1: Changes at each level

	Pre A1 Starters	A1 Movers	Pre A1 Starters and A1 Movers	A2 Flyers	Pre A1 Starters, A1 Movers and A2 Flyers
Pre-revision	466	358	824	471	1,295
Post-revision	508	405	913	519	1,432
Percentage increase	9%	13%	11%	10%	11%

Table 6.2: Total number of words pre- and post-revision

CONCLUSION

The size of the vocabulary list has increased overall by 137 words, an increase of 11%, which adds emerging common words while remaining in line with vocabulary size expectations at CEFR pre-A1 to A2 levels (Milton and Alexiou 2009). Of the additions, 11 words are new names. The majority of new words, however, were added for the following reasons:

1. To introduce words which have recently become high-frequency words.
2. To complete existing lexical sets.
3. To reflect words which were being used naturally by this age group of English language learners.

Examples of additions for Pre A1 Starters include:

1. *Tablet* (recent high-frequency word).
2. *Skateboard* and *tennis racket*, *baseball cap* and *shorts*, *painting*, *poster* and *crayons* to complete the sport, clothes and school-based lexical sets.
3. *Fantastic*, *Hi*, and *Cool!*. Words which we saw from the research were part of candidate vocabulary at this level.

At A1 Movers, examples include:

1. *E-book*, *app* and *laptop*.
2. *Ill* and *sick* (health), *dentist* and *cook* (jobs), *building* and *car park* (places).
3. *Brilliant!*, *little* and *shape*.

And at A2 Flyers:

1. *Chat*, (*open and close*) *a file*, *festival* and *wifi*.
2. *Beetle* and *eagle* (animals), *cereal*, *olive* and *strawberry* (food).
3. *By myself*, *motorway* and *spaceship*.

Words which were removed include *John* (lack of orthographical transparency), *Christmas* and *supper* (culture specific), *fireman/woman* (gender specific) and *CD* and *DVD player* (falling out of daily use).

The revised vocabulary list, which is arranged alphabetically, thematically and grammatically, can be downloaded at cambridgeenglish.org/images/young-learners-handbook-2018.pdf.

Appendix

Note that some words were not new to Pre A1 Starters, A1 Movers and A2 Flyers, but moved between Pre A1 Starters, A1 Movers and A2 Flyers. In the lists below, (*Movers*) indicates words that have been moved from the pre-revision A1 Movers vocabulary list. (*Flyers*) indicates words that have been moved from the pre-revision A2 Flyers vocabulary list.

Added		Removed	
Alice (<i>proper n</i>)	Fantastic! (<i>excl</i>)	pet (<i>n</i>) (<i>Movers</i>)	Bye (-bye) to Bye! (<i>excl</i>)
baseball cap (<i>n</i>)	favourite (<i>adj</i>)	pie (<i>n</i>)	dad(dy) to dad (<i>n</i>)
bat (<i>n</i>) (<i>sports equipment</i>) (<i>Movers</i>)	fun (<i>Movers</i>)	polar bear (<i>n</i>)	mum(my) to mum (<i>n</i>)
bear (<i>n</i>) (<i>Movers</i>)	go to bed (<i>v</i>)	poster (<i>n</i>)	supper (<i>n</i>)
bee (<i>n</i>)	go to sleep (<i>v</i>)	rug (<i>n</i>)	test (<i>n + v</i>)
board game (<i>n</i>)	Hi! (<i>excl</i>)	scary (<i>adj</i>)	Tony (<i>proper n</i>)
boots (<i>n</i>)	Hooray! (<i>excl</i>)	ship (<i>n</i>)	
clap (<i>v</i>)	Hugo (<i>proper n</i>)	shorts (<i>n</i>) (<i>Flyers</i>)	
classmates (<i>n</i>)	jellyfish (<i>n</i>)	silly (<i>adj</i>)	
cool (<i>adj</i>)	kids (<i>n</i>)	skateboard (<i>n</i>)	
Cool! (<i>excl</i>)	kiwi (<i>n</i>)	skateboarding (<i>n</i>)	
count (<i>v</i>)	Mark (<i>proper n</i>)	tablet (<i>n</i>)	
crayons (<i>n</i>)	Matt (<i>proper n</i>)	teddy (bear) (<i>n</i>)	
donkey (<i>n</i>)	meatballs (<i>n</i>)	tennis racket (<i>n</i>)	
Eva (<i>proper n</i>)	painting (<i>n</i>)	thing (<i>n</i>) (<i>Movers</i>)	
fantastic (<i>adj</i>)	paper (<i>n</i>) (<i>Flyers</i>)	zebra (<i>n</i>)	

Table 1: Summary of words added, moved or removed by level: Pre A1 Starters

Added		Removed	
along (<i>prep</i>)	funfair (<i>n</i>)	player (<i>n</i>) (<i>Flyers</i>)	bank (<i>n</i>)
app (<i>n</i>)	goal (<i>n</i>) (<i>Flyers</i>)	pop star (<i>n</i>)	CD player (<i>n</i>)
around (<i>prep</i>)	grow (<i>v</i>) (<i>Flyers</i>)	practice (<i>n</i>)	DVD player (<i>n</i>)
asleep (<i>adj</i>)	helmet (<i>n</i>)	practise (<i>v</i>)	fan (<i>n</i>)
brave (<i>adj</i>) (<i>Flyers</i>)	huge (<i>adj</i>)	roller skates (<i>n</i>)	grandchild(ren) (<i>n</i>)
brilliant (<i>adj</i>)	ice (<i>n</i>) (<i>Flyers</i>)	roller skating (<i>n</i>)	John (<i>proper n</i>)
Brilliant! (<i>excl</i>)	ice skates (<i>n</i>)	sauce (<i>n</i>)	skates (<i>n</i>)
build (<i>v</i>)	ice skating (<i>n</i>)	score (<i>v</i>) (<i>Flyers</i>)	
building (<i>n</i>) (<i>Flyers</i>)	ill (<i>adj</i>)	send (<i>v</i>) (<i>Flyers</i>)	
car park (<i>n</i>)	Julia (<i>proper n</i>)	shape (<i>n</i>)	
circus (<i>n</i>) (<i>Flyers</i>)	laptop (<i>n</i>)	sick (<i>adj</i>)	
Clare (<i>proper n</i>)	little (<i>adj</i>) (<i>Flyers</i>)	sky (<i>n</i>) (<i>Flyers</i>)	
cook (<i>n</i>) (<i>Flyers</i>)	machine (<i>n</i>)	snail (<i>n</i>)	
dance (<i>n</i>)	milkshake (<i>n</i>)	swimsuit (<i>n</i>)	
dangerous (<i>adj</i>) (<i>Flyers</i>)	model (<i>n</i>)	teach (<i>v</i>) (<i>Flyers</i>)	
dentist (<i>n</i>) (<i>Flyers</i>)	net (<i>n</i>)	toothpaste (<i>n</i>)	
e-book (<i>n</i>)	noodles (<i>n</i>)	tractor (<i>n</i>)	
feed (<i>v</i>)	o'clock (<i>adj</i>) (<i>Flyers</i>)	water (<i>v</i>)	
film star (<i>n</i>)	pancake (<i>n</i>)	wave (<i>n</i>)	
fix (<i>v</i>)	penguin (<i>n</i>)	Zoe (<i>proper n</i>)	

Table 2: Summary of words added, moved or removed by level: A1 Movers

Added			Removed
(open and close) a file (n)	enter (a competition) (v)	project (n)	advice (n)
act (v)	explore (v)	puzzle (n)	bright (adj)
amazing (adj)	festival (n)	pyjamas (n)	centimetre (n)
Amazing! (excl)	fire fighter (n)	quiz (n)	Christmas (proper n)
appear (v)	Frank (proper n)	racing (car; bike) (adj)	describe (v)
as (adv)	frightening (adj)	rock music (n)	exam (n)
as ... as (adv)	Go away! (excl)	search (n + v)	fact (n)
at the moment (adv)	gym (n)	shampoo (n)	fireman/woman (n)
bandage (n)	hole (n)	skyscraper (n)	footballer (n)
bank (n) (Movers)	If you want! (excl)	snowboard (n)	get married (v)
beetle (n)	In a minute! (excl)	snowboarding (n)	headteacher (n)
borrow (v)	instead (adv)	Sophia (proper n)	herself (pron)
bracelet (n)	interested (adj)	sore (adj)	himself (pron)
by myself (adv)	invent (v)	spaceship (n)	itself (pron)
by yourself (adv)	invitation (n)	stadium (n)	metre (n)
cereal (n)	knee (n)	step (n)	myself (pron)
chat (v)	land (v)	stone (n)	normal (adj)
costume (n)	manager (n)	strawberry (n)	painter (n)
creature (n)	motorway (n)	stream (n)	partner (n)
cushion (n)	nest (n)	surprise (n)	policeman/woman (n)
cycle (v)	No problem! (excl)	tortoise (n)	secretary (n)
deep (adj)	ocean (n)	touch (v)	single (adj)
delicious (adj)	Oliver (proper n)	trainers (n)	steal (v)
design (n + v)	olives (n)	tune (n)	tape recorder (v)
designer (n)	oven (n)	tyre (n)	tights (n)
disappear (v)	platform (n)	unkind (adj)	toilet (n)
eagle (n)	pleased (adj)	while (conj)	
Earth (proper n)	police officer (n)	wifi (n)	
elbow (n)	pond (n)	x-ray (n)	
engine (n)	pop music (n)	yoghurt (n)	
enormous (adj)	program (n)		

Table 3: Summary of words added, moved or removed by level: A2 Flyers

References and further reading

- Cambridge English Language Assessment (2016) *Principles of Good Practice: Research and Innovation in Language Learning and Assessment*, Cambridge: Cambridge English Language Assessment.
- Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge: Cambridge University Press.
- Elliott, M and Stevenson, L (2015) Grading and test equating, *Research Notes* 59, 14–20.
- Field, J (2011) Cognitive validity, in Taylor, L (Ed) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, Studies in Language Testing volume 30, Cambridge: UCLES/Cambridge University Press, 65–111.
- Field, J (2013) Cognitive validity, in Geranpayeh, A and Taylor, L (Ed) *Examining Listening: Research and Practice in Assessing Second Language Listening*, Studies in Language Testing volume 35, Cambridge: UCLES/Cambridge University Press, 77–151.
- Field, J (forthcoming 2017) The cognitive validity of reading and writing tests designed for young learners, in Papp, S, and Rixon, S, *Examining Young Learners: Research and Practice in Assessing the English of School-age Learners*, Studies in Language Testing volume 47, Cambridge: UCLES/Cambridge University Press.
- Fulcher, G, Davidson, F and Kemp, J (2011) Effective rating scale development for speaking tests: Performance decision trees, *Language Testing* 28 (1), 5–29.
- Hattie, J and Timperley, H (2007) The power of feedback, *Review of Educational Research* 77 (1), 81–112.
- Jones, N and Saville, N (2016) *Learning Oriented Assessment: A Systemic Approach*, Studies in Language Testing volume 45, Cambridge: UCLES/Cambridge University Press.
- McNamara, T F (1996) *Measuring Second Language Performance*, London: Longman.
- Milton, J and Alexiou, T (2009) Vocabulary size and the Common European Framework of Reference for Languages, in Richards, B, Daller, M H, Malvern, D, Meara, P, Milton, J and Treffers-Daller, J (Eds) *Vocabulary Studies in First and Second Language Acquisition: The Interface Between Theory and Application*, Basingstoke: Palgrave Macmillan, 194–211.
- Moeller, A J, Creswell, J W and Saville, N (Eds) (2016) *Second Language Assessment and Mixed Methods Research*, Studies in Language Testing volume 43, Cambridge: UCLES/Cambridge University Press.
- Rasch, G (1960/1980) *Probabilistic Models For Some Intelligence And Attainment Tests*, Copenhagen: Danish Institute for Educational Research/Chicago: The University of Chicago Press.
- Upshur, J A and Turner, C E (1995) Constructing rating scales for second language tests, *ELT Journal* 49, 3–12.
- Upshur, J A and Turner, C E (1999) Systematic effects in the rating of second language speaking ability: Test method and learner discourse, *Language Testing* 16 (1), 82–111.
- Weir, C J (2005) *Language Testing and Validation: An Evidence-based Approach*, Basingstoke: Palgrave Macmillan.

Authors



Sarah Albrecht manages the Pre A1 Starters, A1 Movers and A2 Flyers Listening component. She has a PGCE in teaching languages at primary level and is embarking on an MSc in Digital Education. She previously worked as a teacher trainer and educational advisor, supporting English and Content and Language Integrated Learning (CLIL) teachers for the Madrid Regional Government.



Chris Cullen manages the Pre A1 Starters, A1 Movers and A2 Flyers speaking component. He has an MA in Applied Linguistics. He previously established companies that specialised in teaching Business English to a diverse range of clients, such as the Portuguese Central Bank in Lisbon and China Shipping in Valencia.



Kathryn Davies manages the Pre A1 Starters, A1 Movers and A2 Flyers Reading/Writing component. Prior to this, she edited coursebooks for secondary schools at Cambridge University Press. Kathryn has taught in the UK, Japan and Spain. She is Delta qualified, has an MA in Linguistics and is currently pursuing a Postgraduate Certificate in Educational Assessment.



Maggie Dunlop is responsible for research related to the Pre A1 Starters, A1 Movers and A2 Flyers exams. Maggie taught EFL for five years, and holds a PhD in Second Language Education, specialising in second language assessment. Her interests lie in using evidence to support language learning.



Mark Elliott is responsible for grading and test equating for the Pre A1 Starters, A1 Movers and A2 Flyers exams. Mark taught EFL and worked in materials development before joining Cambridge Assessment English, where he has worked in the Assessment, Research and Validation, and Validation and Data Services departments. Mark holds an MA (Cantab) in Mathematics and an MA in ELT and Applied Linguistics.



Lynne Stevenson manages the team which produces the Pre A1 Starters, A1 Movers and A2 Flyers exams. Prior to joining Cambridge Assessment English, Lynne worked in publishing and then taught EFL for several years. Her interests include learning-oriented assessment for young learners. She holds an MA in Applied Linguistics and TESOL.

Contact us

Cambridge Assessment English
1 Hills Road
Cambridge
CB1 2EU
United Kingdom

All details are correct as of January 2018.



 cambridgeenglish.org/younglearners

 [/CambridgeEnglish](https://www.facebook.com/CambridgeEnglish)

 [/CambridgeEnglishTV](https://www.youtube.com/CambridgeEnglishTV)

 [/CambridgeEng](https://twitter.com/CambridgeEng)

 [/cambridgeenglish](https://www.instagram.com/cambridgeenglish)