

Task types for language assessment

To accompany Chapters 5 and 6 of *Exploring Language Assessment and Testing*.

The following tables are indicative of the variety of task design options that are open to the language assessment developer. Assessment designers should find them a helpful reference when designing tasks to assess different areas of language knowledge, skills or abilities. They range from the relatively mechanical manipulation of forms and patterns to complex performance tasks.

The starting point in designing assessments is to think systematically about language use in the target language use (TLU) domain. The objective is usually to create assessment tasks that emulate TLU tasks so that performance on the assessment reflects, to the extent possible, target language use. Do assessees have to carry out similar activities, employing similar processes both in the assessment and in the TLU domain?

To achieve this objective, the assessment designer can use these tables to find appropriate combinations of elements – a selected response gap fill task could involve choosing a word to fill a gap in a sentence; a similar constructed response gap fill task would not offer any words to choose from, but would simply require the assessee to write the word needed to fill the gap.

Separate tables are presented, distinguishing between:

- receptive skills and productive skills (subdivided into writing and speaking)
- response formats (selected response, short constructed response, extended response, personal response)
- task types (gap-fill, sequencing, summarising etc.)

The categories presented are not exclusive. Gap filling might also involve summarising. Listening tasks might also involve some reading and writing or speaking. The listed elements may be adapted and combined to provide an infinite range of assessment possibilities.

In each table, there are four columns.


Task type/Response format: A name for the relevant task type or response format.

Description: A brief description with simple illustrative examples.

What it does it assess? An outline of the knowledge, skills or abilities that each task/ format is most often used to assess.

Comments: Additional commentary covering issues such as key advantages/ disadvantages and administrative practicalities.

Receptive skills: Response Formats

Selected response	Description	What does it assess?	Comments
True/False	<p>Assesseees decide whether each of a number of statements is true (is supported by the text/ recording), or false (is contrary to the text/ recording). Sometimes a third option - not stated (no information in the text/ recording) – is included.</p>  <p>There are some pencils next to a cup. Yes <input type="radio"/> No <input checked="" type="radio"/></p>	<p>True/False items are often used to assess the explicit meaning of individual sentences or idea units in spoken language, but can also be used in other ways: for example in judging the accuracy of summaries of opinions expressed in a text.</p>	<p>An obvious weakness of True/False items – one that is especially problematic for more summative assessment purposes – is that assesseees have a 50-50 chance of guessing the right answer. This means that only scores higher than 50% can be meaningful. Although different scoring strategies have been tried to penalize assesseees who guess, none of them has proved to be satisfactory. Adding a third ‘not stated’ or ‘no information’ option helps reduce the size of the score attainable through chance, but also makes items more difficult to write. The difference between ‘no information’ and implicitly ‘false’ is often fuzzy. The ‘no information’ option is particularly questionable – and should be avoided – in tests of listening. Assesseees are not able to go back through a recording to confirm whether or not any relevant information is given.</p>
Multiple Choice	<p>Assesseees choose the correct answer to each item from among a set of options (three is widely considered the most appropriate number) to answer a question or finish an incomplete statement.</p> <p>How many pencils are in the cup?</p> <p>a) <input checked="" type="radio"/> three b) <input type="radio"/> four c) <input type="radio"/> five</p>	<p>The multiple choice item format is extremely flexible and can be used to assess the integration of ideas across texts as well as vocabulary, grammar and propositional meaning. However, creating effective items requires great skill on the part of the item writer – as well as careful editing and review and trialling of the material.</p>	<p>Multiple choice questions provoke strong reactions. In some countries they are almost the only format that is used in tests. In other countries they are viewed with great suspicion and are very rarely used. Like True/False they can be susceptible to guessing. Although this means that scores lower than 33% (for tests made up of three-option items) cannot be meaningful, provided there are reasonable numbers of items, scores above this level are unlikely to occur by chance. Having four or more options may reduce the odds of guessing the correct answer, but there is rarely a plausible fourth option. This means that test writers often waste a good deal of time inventing fourth options that most assesseees can immediately discount – leaving them with a three-way guess.</p>

Selected response	Description	What does it assess?	Comments
Matching	<p>Matching is similar to multiple choice, but each option can be matched to one of several stems. Examples include</p> <ul style="list-style-type: none"> - matching words or sentences to the pictures they describe - matching words to definitions - matching sentence fragments or matching responses to their questions - matching dialogues or texts to likely contexts - choosing the most suitable headings or summary statements that match paragraphs or sections of a text 	See multiple choice, above.	<p>Matching is a very versatile task type with a lot of possible variations. Usually each option is matched to just one stem, but matches between one stem and multiple options or multiple options and one stem are also possible. Whichever approach is adopted, the matching process must be made completely clear to the assesseees in the instructions.</p> <p>In test situations, it may be best to limit the format to one to one matching as this is the least likely to cause confusion.</p> <p>The risk of scoring more than a few points through blind guessing is reduced relative to multiple choice, depending on how many options are given.</p> <p>It is usual and sensible to include a few more options than there are stems. One or more of the options will not correspond to any of the stems. This means that the task will not become a puzzle in which assesseees who know four of five answers can guess the fifth through a process of elimination.</p>

Constructed response	Description	What does it assess?	Comments
Short Response	<p>Assesseees give a brief response of one word or a short phrase. Responses may be written or spoken.</p> <p>Q. Where are the pencils? A. <i>In the cup.</i></p>	Short responses are better suited to assessing grammar, vocabulary and simple propositional meaning than higher-level processes	Short constructed response items do not present the same issues of guessability as selected response items (questions that can be answered yes or no or that offer ‘which one?’ choices are best avoided). They are also a little easier to construct because distractors are not needed. However, assesseees may sometimes guess the answers to comprehension questions by copying strings of words from the text and particular care must be taken to rule out this possibility. As short response items are not as controlled as selected response, care must be taken through reviewing and trialling the test materials to ensure that the scoring key is comprehensive and fair to all.
Extended Response	<p>Assesseees respond at greater length. Responses may range in length from a clause or sentence upwards.</p> <p><i>Input.</i> There was no rain for six months. The grass was dry and brown. The cattle did not produce any milk. None of the crops in the fields would grow.</p> <p>Q. Why didn’t the crops grow? A. <i>Because there was no rain for six months</i></p>	Extended responses are well suited to assessing inference, text level and intertextual relationships as well as basic propositional comprehension.	Extended response items only become available as indicators of reading or listening comprehension when it can be assumed that the assesseees have sufficient writing or speaking ability to give longer responses. Effective extended responses not only require comprehension, but also involve writing or speaking abilities. It may be appropriate to consider tasks that require extended responses as assessments of ‘reading-into-writing’, ‘listening-into-writing’ and so on (see Chapter 6 for ways of scoring these).
Personal Responses	<p>Personal response formats are not distinguished by the extent of the response (they may also be <i>selected</i>, <i>short answer</i> or <i>extended</i>). Their characteristic feature is that they record the assessee’s personal reaction to the input.</p> <p>They fall into two basic categories: <i>affective responses</i> – concerned with the assessee’s feelings about the input material and <i>self-assessment</i> – the assessee judges how well he/she understands the input.</p> <p><i>Affective response</i> Selecting reading or listening material, questionnaires and self reports about personal reaction to input material.</p> <p><i>Self-assessment</i> Using a rating scale to reflect on quality of own performance, strengths and weaknesses. Setting realistic learning targets.</p>	<p><i>Affective response</i> Motivation and interest.</p> <p><i>Self-assessment</i> Learner self-regulation: setting and adjusting goals, monitoring progress.</p>	<p><i>Affective response</i> Motivation is of great importance in language learning and teachers need to gain a clear picture of the kinds of input that encourage learners to want to read or listen. Providing some choice and helping learners to select suitable and interesting input material can help to foster motivation.</p> <p><i>Self-assessment</i> It is beneficial to give learners of all ages a degree of responsibility for their own learning, help them to set goals, recognize the gap between their current performance and their goal and to identify steps they can take to improve. Regular self-assessment activities and effective student-teacher communication can promote this kind of reflective learning.</p>

Receptive skills: Task Types

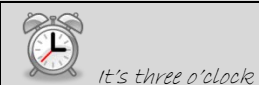
Task Type	Description	What does it assess?	Comments
Gap-fill	<p>Assesseees write, speak or select words to fill gaps in a text (or recording) – variations include</p> <ul style="list-style-type: none"> - short answer gap-fill: productive responses, assesseees write the words that fill the gaps. - multiple choice gap-fill: selecting from a bank of options provided for each gap - matching gap-fill: choosing from a pool of options, any option from the bank could be chosen to fit any gap. <div> <div>There is a cup ____ a table.</div> <div>. on .</div> </div> <div> <div>There are ____ pencils in the cup.</div> <div>. three .</div> </div> <p>In variations on the gap fill item type, assesseees may insert longer stretches of language – phrases, sentences or paragraphs – into the gaps.</p>	<p>Mainly grammatical and vocabulary knowledge.</p> <p>Gap fill items are the mainstay of traditional grammar and vocabulary tests.</p>	<p>Gaps should generally be indicated with lines or sounds of equal length (unless the intention is to give a clue about the length of the missing word). Sometimes the gaps are not indicated at all: assesseees have to both locate the gaps and fill them, but this is demanding. Hints (such as the first letter of a missing word) can be used to help the assessee and constrain the range of possible responses.</p> <p>In listening, gaps may be indicated by a sound that replaces a missing word, phrase or contribution to a conversation. Sometimes a wide range of different words might reasonably be used to fill the gap – some of these may not be obvious to the test writer. Care must be taken to prepare a comprehensive scoring key.</p>
Information transfer	<p>Assesseees transfer information from a text/ recording into a non-verbal form by completing a diagram, table, grid etc. or by tracing a route on a map.</p> <div> <div>Label the picture with these words:</div> <div> <div>'cup'</div> <div>'pencils'</div> </div> </div> <div> <div>Draw three pencils in a cup.</div> </div>	<p>Vocabulary knowledge. Detailed propositional comprehension.</p>	<p>Tasks of this kind reduce the role of written language and can appear reasonably realistic. However, they sometimes demand quite high levels of visual awareness and assesseees who are unfamiliar with them can struggle to understand what is required.</p>
Intrusion	<p>Assesseees read a text/ listen to a recording and identify any words (or phrases, sentences, paragraphs) which do not belong.</p> <div> <div>There for were three pencils in a cup.</div> <div>for</div> </div> <div> <div>One day the cup fell to the floor. It was -</div> <div>-</div> </div> <div> <div>smashed box into small pieces.</div> <div>box</div> </div>	<p>Mainly vocabulary and grammatical knowledge.</p>	<p>A convenient way of presenting written tasks of this kind is to require a judgement for each line of a text. If there is an intrusion in the line, the word is written in the space beside that line. If not, the space is left blank. One point is awarded for each correct judgement.</p> <p>This kind of task can appear rather puzzle-like and so is often unpopular with assesseees.</p>

Task Type	Description	What does it assess?	Comments							
Sequencing	<p>Assessee re-order elements to build sentences, paragraphs or texts.</p> <div><p>Put these sentences in the correct order.</p><p>The Sears Tower</p><p>A Built in 1974, it is the most famous building in the city.</p><p>B Visitors can take a lift to the observation deck on the top floor.</p><p>C Chicago, with a population of 3 million people, is America’s third largest city.</p><p>D Right in the middle of the city you can see the Sears Tower.</p><p>E They can even look straight down through a glass floor to see people walking in the street below.</p><p>F From here they get some great views of the city and the Chicago river.</p><p>G It is also the tallest, being 442 metres high.</p><table><tr><td>C</td><td>D</td><td>A</td><td>G</td><td>B</td><td>E</td><td>F</td></tr></table></div>	C	D	A	G	B	E	F	Grammatical knowledge, discourse level knowledge, relationships between the elements of a text.	Ordering sentences in an argument or narrative text should require knowledge of cohesion and connections between the sentences. It can be challenging to create tasks in which only one order is unambiguously correct. Scoring sequencing tasks can be complex. The items depend to some extent on each other. Sometimes such items are treated as dichotomous (one point for a fully correct sequence), but it is more usual to use a polytomous (partial credit) scoring system, awarding one point whenever two elements are correctly ordered. In the example one point is scored for C-D, one for D-A and so on. This means that it is impossible to score 2 out of three (or six out of seven etc.) on this kind of task.
C	D	A	G	B	E	F				
Summary and recall	<p>Assessee summarize one or more texts/ recordings by choosing summary statements or by constructing a summary in their own words.</p> <p>There are many variations including selected and constructed response formats.</p> <p>Extended response alternatives include</p> <ul style="list-style-type: none">different word length constraints – e.g. a 50, 35 or 10 word summary of a 100 word text;summary recallwriting or relating a summary without being able to see the input text; paraphrase – a summary ‘in your own words’ that avoids repeating words or phrases from the inputcomparative summaries that synthesize ideas from different texts. <p>Selected response options include</p> <ul style="list-style-type: none">judging the accuracy of summaries (True/ False or multiple choice)matching summaries to stretches of textselecting and ordering summary statements <p>In gap-fill summaries, assessee complete a partial summary of the text/ recording using their own words, words from the text, or words/phrases from a pool of options to fill the gaps.</p>	<p>Overall understanding of a complete text/ recording (or section of text/ recording).</p> <p>However, composing successful summaries requires sophisticated writing skills (or speaking skills if it is an oral summary) as well as comprehension of the input.</p>	<p>Effective paraphrase and synthesis are essential in some target language use settings such as academic study. In other settings, simpler recall and retelling may be more appropriate.</p> <p>Gapped summaries are easier to score than complete summaries and require less writing (or speaking skill), but care must be taken that responses are not suggested to the assessee by the grammar and vocabulary of the gapped text.</p>							
Dictation	<p>Assessee write (or select) words, phrases or sentences as they hear them.</p> <p>In <i>complete dictations</i>, the recording is regularly paused at phrase or sentence boundaries and the assessee write all the words that they hear.</p> <p>In <i>partial dictations</i>, they write or choose words to fill in gaps in the transcript or complete a form or table.</p>	Mainly word recognition and grammatical knowledge.	Like gapped summaries, gapped dictations are easier to score than the complete alternative and can be focused on specific issues for diagnostic purposes.							

Task Type	Description	What does it assess?	Comments
Repeat	<p>Repeats are similar to short dictations, but instead of writing the words they hear, the assessee repeat them back orally.</p> <p>A correct response is one that accurately repeats the prompt. Scores may be dichotomous (right/wrong) or partial credit (e.g. 0, 1, 2 or 3 points according to how closely the response matches the prompt)</p>	Mainly word recognition and grammatical knowledge.	Having learners repeat short strings of words can provide diagnostic insights into difficulties with (receptive) word recognition and (productive) articulation. However, it may not be immediately clear which of these is causing problems for the assessee, or whether the problems are attributable to short term memory issues.
Read aloud	Assessee reads a word/ phrase/ sentence or text aloud. If the response accurately reflects the prompt, it is considered correct. Alternatively, reading aloud may be scored using a checklist or rating scale.	The relationship between words on the page or screen and the sound system of the language.	Perhaps most appropriate with beginning readers who are learning about word-sound correspondences. Both reading and speaking skills are involved. Fluent reading aloud does not equate to comprehension.
Rating tasks	<p>Paralleling the division of personal response types (see above) into self-assessment responses and affective responses, tasks for learner reflection can be divided into <i>rating tasks</i> and <i>recording tasks</i>.</p> <p>Rating tasks involve making judgements. They encompass a wide range of activities aimed at encouraging learners to take responsibility for their learning. Assessee may use checklists, rating scales or commentaries to set reading and listening goals and to reflect on their level of comprehension.</p> <p>Examples include:</p> <ul style="list-style-type: none"> - assessee highlight sections of text that they found difficult or circle a number representing the percentage of a text/ recording they felt they understood - assessee use a checklist to report what strategies they used when reading a text/ listening to a recording or identify strategies to try next time they read/ listen - assessee use a rating scale to assess their own reading/listening level and to consider what would need to change in their reading/ listening for them to reach the next level on the scale 	All levels of comprehension. Aspects of motivation and feelings towards reading/ listening.	<p>Outcomes of self-assessment can be discussed in class, or in student-teacher conferences.</p> <p>Self-assessment is not at all easy and learners usually need extensive practice and support before they are able to benefit fully. It is most effective when carried out frequently as a regular feature of the classroom.</p>
Recording tasks	<p>Recording tasks involve documenting what has been done.</p> <p>Examples include reading/ listening logs and portfolios (but note that these may involve rating as well as recording).</p> <p>In reading/ listening logs, assessee keep an ongoing record of their reading and listening activities. This may be as part of an extensive reading/ listening programme. For each entry, the learners may complete a brief form with details of, for example, what they have read/ listened to – or viewed (What did you read/ listen to?) and a brief summary of the content (What was it about?). The recording aspect may be combined with affective rating of the content (How much did you enjoy the book/ movie?) and self-assessment (How much did you understand?)</p>	Focus is on logging what the assessee has done. This is often a preliminary to affective evaluation or self-assessment. Grading and correction of language are not usually involved.	<p>The objective of this kind of activity is to discover more about what motivates assessee to read or listen to the foreign language: what they enjoy or dislike and what interests them.</p> <p>This information can be used to encourage them to do more reading and listening beyond the classroom.</p>

Writing Task Types

Task Type	Description	What does it assess?	Comments
Copying (Repeats)	Copying tasks involve accurately copying letters, symbols, words or phrases in the target language. The spoken variant of this task type is 'Repeats' (see above).	Orthography: formation of letters and words. Short term verbal memory, pronunciation.	Copying is generally of limited value as an assessment tool beyond the earliest stages of literacy in the target language – and may be redundant when the assessee's native language shares a script with the target language. Repeating back words or phrases can help in identifying difficulties with pronunciation.
Sequencing	Assessee reorders (spoken or written) sentence fragments to build sentences. <div> <div>six eggs/ in a box/ there are</div> </div>	Grammatical and vocabulary knowledge. The focus of sequencing tasks is on the form of the sentence/ speech unit or paragraph. There is little scope for any communication.	Sequencing is also listed as a receptive task type. In tests of production, the focus is more often on local organisation within the sentence rather than on connections between sentences or paragraphs.
Transformations	Assessee transforms (spoken or written) base words into the appropriate (spoken or written) form, recasts sentences so that they have a different form, but equivalent meaning or combines simple sentences to form more complex sentences. <div> <div>Have you [<i>buy</i>] _____ any eggs today? <i>bought</i></div> <div> <div>The piano was so heavy that three men were needed to lift it.</div> <div> <div>It took _____ because it was so heavy.</div> <div><i>three men to lift the piano</i></div> </div> </div> </div>	Grammatical and vocabulary knowledge. These item formats can provide a reasonable indication of formal grammatical knowledge. Although more versatile than sequencing, they provide little evidence that the assessee is aware of the meaning of the forms or of the communicative effects of ordering information in different ways.	In the classroom, tasks of this kind can provide teachers and learners with helpful insights into how well aspects of the language system have been understood. However, the ability to manipulate a form in a supported context should not be mistaken for the ability to use it accurately and appropriately in spontaneous production.

Task Type	Description	What does it assess?	Comments
Error correction	<p>Assessee identifies the parts of a text or short recording that contain errors. Variations include providing hints on the locations of the errors (as in this example) and requiring the assessee to make the necessary corrections.</p> <p>Assessee should be given guidance on the number of errors to locate so that they know when the task is completed.</p> <div> <p>Find the error in this sentence.</p> <p>Q: Protein (a) <u>provides</u> the body with (b) <u>roughly</u> 10 to 15 per cent of (c) <u>its</u> dietary energy, and is (d) <u>needing</u> for growth and repair.</p> <p>A: d</p> </div>	Grammatical and vocabulary knowledge.	<p>This task type can be based on learners' own production to raise awareness of common errors and encourage self- and peer-correction.</p> <p>Items are relatively easy to construct and score.</p> <p>Items that are limited to identifying errors do not indicate that the assessee is capable of correcting the error – and error correction is in any case only a very limited part of writing ability. Items that require the assessee to make corrections are more open-ended and therefore harder to score. A comprehensive scoring key is needed.</p> <p>The error identification format is, like other selected response types, vulnerable to guessing so that scores below the level of chance (i.e. 33% or less on test made up of three option multiple-choice items) cannot be interpreted.</p>
Sentence building	<p>Based on isolated words, jumbled sentences, pictures or question prompts, assessee writes or says individual sentences.</p> <div>  </div> <div> <p>Write a sentence using these words: 'fortunate' 'realise' 'bonus'.</p> </div> <div> <p>You are eating lunch in the student canteen and you accidentally spill your drink onto the young man sitting next to you. You have never met him before.</p> <p>What do you say?</p> </div>	Grammar/vocabulary knowledge, pronunciation, sentence level organisation. Can be adapted to address basic pragmatic awareness.	<p>Sentence building tasks risk becoming transformation exercises: transforming questions or prompts to make statements. The use of pictures and symbols rather than text can help to limit this kind of support.</p> <p>This kind of task is often decontextualised. As a result, the responses may not give the user much evidence of whether assessee knows when and where it is appropriate to use the assessed language forms.</p> <p>It can be surprisingly challenging to assess knowledge of specific grammar or vocabulary knowledge without giving over-complicated instructions, giving away unintended clues to the correct answer or using metalanguage which assessee may not know – like <i>verb tense</i> or <i>adjective</i>. The intention in language assessment is generally to assess the ability to use the language rather than the ability to describe it.</p> <p>Careful and critical reviews by colleagues and trialling of material are needed to minimise such problems.</p>

Extended Writing Task Types

Task Type	Description	What does it assess?	Comments				
Guided writing	<p>This is similar to sentence building, but the prompts are sequenced in such a way that the responses are thematically linked together to form a coherent text.</p> <p>The structure of the response or interaction is provided by the input.</p> <p>Examples include:</p> <p>Form filling</p> <table><tr><td>Name:</td></tr><tr><td>Favourite sport:</td></tr><tr><td>Why do you enjoy this sport?</td></tr></table> <p>Writing from notes, outlines or guiding questions.</p> <table><tr><td>Write about your favourite sport. What is the sport? How often do you play it? Where do you play it? Who do you play it with? etc.</td></tr></table>	Name:	Favourite sport:	Why do you enjoy this sport?	Write about your favourite sport. What is the sport? How often do you play it? Where do you play it? Who do you play it with? etc.	<p>Vocabulary, grammar, sentence level organisation.</p> <p>Text level organisation is not extensively tested as it is supported by the input.</p>	<p>As in sentence building, prompts may provide a good deal of the language that the assessee needs, reducing the task to a series of sentence transformations.</p> <p>The amount of guidance can be adjusted to suit the level of the assessees. As they develop their writing abilities, guidance can be phased out.</p> <p>A semblance of authenticity can be added by embedding the guidance in a form, letter or set of instructions to which the assessee must respond.</p> <p>Responses to tightly guided compositions are more easily compared than responses to more open-ended tasks and may be scored reliably by raters with minimal training.</p>
Name:							
Favourite sport:							
Why do you enjoy this sport?							
Write about your favourite sport. What is the sport? How often do you play it? Where do you play it? Who do you play it with? etc.							
Information transfer	<p>These supported writing tasks are the mirror image of the information transfer task type described in the Reading/ Listening section above.</p> <p>Assesseees work from data in a picture, map, table or diagram and transfer the information into written form. Examples include descriptions, picture stories, directions to a location, instructions based on a sequence of diagrams, a report based on data in a graph etc.</p> <p>Responses can range from quite basic sentence building up to extended reports.</p> <p>A variation involves providing input in the learners’ first language. The assesseees do not translate this, but summarise or report on the input text/ recording.</p>	<p>Grammar, vocabulary, textual organisation, coherence.</p> <p>The cognitive demands of these tasks can be increased by adding an interpretive element requiring assesseees to go beyond reporting the data to interpreting and evaluating what they are given: possible causes and consequences, alternative interpretations etc.</p>	<p>As with guided writing, the nature of the responses can be very strictly or more loosely controlled: the more controlled the response, the easier it is to award scores.</p> <p>Providing data to be used in the response has the advantage that, with clear written instructions, the task requirements can be made very clear to the assesseees.</p> <p>The response requires little imagination or creativity, giving uninspired assesseees an equal chance to perform, but making the process more mechanical and less creative.</p>				

Task Type	Description	What does it assess?	Comments
Narratives, reports, essays	<p>Assesseees are asked to write a set number of words in a specified genre (story, letter, report, essay etc.) on a given topic that is of relevance to their interests (or to the TLU domain)</p> <p>A degree of contextualisation is normally included in the instructions. These may specify</p> <ul style="list-style-type: none"> - An addressee – writing to a friend, a customer, a senior colleague, a teacher, etc. - A setting – writing for work, school, personal expression etc. - A rhetorical task or tasks – writing to inform, complain, persuade etc. 	<p>Sentence level, discourse level organisation, pragmatic and sociolinguistic awareness.</p> <p>The demands of the task can be adjusted through features such as topic familiarity (concrete personal and familiar everyday topics tend to be easier for assesseees than more abstract concepts), rhetorical function (information giving is generally easier than evaluation), opportunities for planning and preparation (more planning time leads to improved performance) and length requirements.</p>	<p>The range of possible tasks is very wide. They can range from concrete descriptions and personal narratives to more evaluative essays and reports. Text types might include academic essays, professional reports, letters.</p> <p>Open-ended tasks like these are very difficult to score reliably — they require rater training and scoring by multiple raters.</p> <p>Contextualisation helps to suggest to the assessee sociolinguistic aspects of the task such as the appropriate level of formality of the response.</p>
Reading (or listening) into writing	<p>Assesseees are given one or more extended texts to read (or listen to recordings) and are prompted to produce a written response that draws on information from the input(s).</p> <p>Reading or listening can also be used as input for extended speaking and interview tasks.</p> <p>Much of the writing that people do in academic and professional settings is based on reading and interpreting the work of other writers. This kind of task reflects the need to integrate ideas from other writers without simply copying their words.</p>	<p>Sentence level, discourse level organisation, paraphrasing/ summarising skills, pragmatic and sociolinguistic awareness.</p>	<p>Direct copying from the input is problematic for scoring. Task instructions need to be very clear on how this is addressed so that assesseees understand the expectations.</p> <p>It may be preferable not to allow assesseees to consult the input text as they write, but to work from notes that are restricted to words or phrases.</p> <p>These tasks reduce the impact of topic familiarity as every assessee is given something to write about.</p> <p>In common with other extended writing tasks, they are difficult to score reliably: they require multiple scoring by skilled and trained raters.</p>

Task Type	Description	What does it assess?	Comments
Multiple drafts	<p>Given an extended writing assignment, assessees submit one or more drafts. On receiving targeted feedback, they try to improve their work before submitting a final version.</p> <p>Variations on this task type can be used in preparing extended spoken monologues.</p>	<p>Sentence level, discourse level organisation, paraphrasing/ summarising skills, pragmatic and sociolinguistic awareness.</p> <p>The key objective of assessing multiple drafts is to gain insights into how well assessees are able to take account of feedback so that their work more closely approximates to good writing.</p>	<p>It has long been recognised that skilled writers generally carefully edit and reorganise their work, going through multiple drafts before releasing a final version.</p> <p>Assessing this process should encourage learners to engage in self-monitoring.</p> <p>Multiple drafting is difficult to operationalise in traditional tests, but is more easily managed in the classroom.</p>
Portfolio Assessment	<p>Over a period of time assessees assemble a range of their written work (or recordings of spoken performances).</p> <p>Variations include 'performance portfolios' – in which learners select their best work for the portfolio – and 'learning portfolios' – in which all work, including drafts – is assembled (Wiliam 2011).</p>	<p>A key advantage of portfolios is that they provide evidence for development over time. They can inform conclusions about how far an assessee has improved in targeted skills.</p> <p>Wiliam (2011) argues in favour of learning portfolios for formative purposes because they more clearly indicate progress. Hap-Lyons and Condon (2000) argue for the benefits of performance portfolios as learners take decisions about which work represents their best performance.</p>	<p>Portfolios can be used for both summative and formative purposes.</p> <p>They encourage self-assessment and reflection by assessees because they provide opportunities for improvement, offer evidence of learning and development over time and, if paired with well-constructed rating scales and teacher-student conferencing, show how performance increasingly approximates to the expected standards.</p> <p>Grading can be delayed until the end of the portfolio period, encouraging learners to improve their work before it is graded.</p>

Speaking Task Types

Task Type	Description	What does it assess?	Comments
Extended monologues	<p>Assesseees are asked to talk for a set length of time on a given topic that is of relevance to their interests (or to the TLU domain).</p> <p>A wide range of tasks falls into this category:</p> <ul style="list-style-type: none"> - Reading aloud - Information transfer (picture stories, descriptions, show and tell) - Guided speech using notes, outlines or guiding questions - Impromptu presentation (unsupported) 	<p>Sentence (or idea unit) level, discourse level organisation, pragmatic and sociolinguistic awareness.</p> <p>As with written stories and essays, the demands of the task can be adjusted through features such as topic familiarity (personal anecdotes tend to be much easier for assesseees to relate to than thoughts on more abstract topics), rhetorical function (information giving is generally easier than evaluation), length requirements, guiding questions or frameworks and opportunities for planning and preparation (more planning time should lead to improved performance).</p>	<p>Extended speaking is more challenging to score than extended writing as it is more difficult to go back and review the performance. For this reason, and to facilitate effective feedback to the speaker, it is important to record what is said.</p> <p>As with extended writing, rater training and scoring by multiple raters are essential when important decisions are at stake.</p> <p>In classroom assessment, monologues can be difficult to manage. Options include having a different learner give a monologue in each class session; having learners give presentations to smaller groups, selectively monitoring these; or asking learners to record their presentations.</p>
Interview	<p>The interviewer (teacher, examiner or fellow assessee) puts a series of questions to each assessee.</p> <p>Pictures or other forms of input may be used to provide input to support the interaction.</p> <p>Roles may be reversed at some points in the interview, with interviewees collectively or individually interviewing the interviewer.</p>	<p>Sentence (or idea unit) level, discourse level organisation, pragmatic and sociolinguistic awareness, interactive language.</p> <p>Questions can be selected to reflect the purpose of the assessment – simple personal questions for beginning learners; hypothetical questions on abstract topics for higher level learners; work-related questions for assessments directed towards professional contexts; questions designed to elicit specific language points for diagnostic, formative purposes.</p> <p>A limitation of the interview is that it may only elicit a relatively limited range of language. As with monologues, pictures, short video clips and texts can be used as a basis for questioning to take the interaction away from basic personal exchanges. However, more transactional functions (offering, purchasing, negotiating, complaining) are rarely required – unless an element of role play is used (see below).</p> <p>An element of rehearsal may be helpful for beginners, but makes the interview less genuinely interactive, limiting it as an assessment of spontaneous speaking ability. If they become the norm, rehearsed interviews can prevent learners from developing the ability to interact more spontaneously.</p>	<p>Interviews are more difficult to score than extended monologues, especially when the assessor is also a participant in the interaction. It is unclear how far the assessee's performance has been affected by the performance of the interviewer – the interview performance (the basis for the score) is really a shared product.</p> <p>Questions may be fixed or invented during the interview, but they should not be known to the assesseees in advance. For comparability, every assessee should be asked the same questions. But this makes the interview predictable. Assesseees cannot all be interviewed at the same time so the first interviewee can tell all the others what to expect, encouraging memorised responses. The solution may be to create a bank of questions of similar difficulty to draw from.</p> <p>One to one interviews with the teacher are difficult to manage in classroom settings. Alternatives include learner-learner interviews (learners question each other using question cards) or interviews using recorded questions in a language laboratory. Both strategies allow the teacher to focus on scoring without having to participate in the interaction, but also mean that the teacher has less control over the questioning process.</p>

Task Type	Description	What does it assess?	Comments
Information Gap	<p>Participants are each given incomplete information. They must interact to combine their knowledge in order to complete a task.</p> <p>Task options include:</p> <ul style="list-style-type: none"> - Picture differences - Table/ picture completion - Reconstructing a story from partial narratives 	<p>Sentence (or idea unit) level, discourse level organisation, pragmatic and sociolinguistic awareness, interactive language.</p> <p>Information gap activities can elicit questioning and give opportunities for both (or all) participants to take a leading role.</p> <p>They can provide very clear goals for the interaction and, unlike interviews, can involve more transactional language.</p> <p>On the other hand, the language tends to be limited to the exchange of concrete information.</p>	<p>Although they can be time-consuming to prepare, and (like other assessments of speaking) difficult to score, information gap activities are quite straightforward to organise in the classroom and are well suited to pair and group interaction. Communicative success is relatively easy for teachers and learners to judge.</p> <p>These tasks are often well-suited to lower level learners as the content can be controlled, much of the language can be repeated and a good deal of support provided to the assessee.</p>
Discussion	<p>Given a topic and perhaps a few guiding questions to consider, assesseees discuss an issue and try to reach a consensus.</p> <p>Topics for discussion can take many forms ranging from deciding what to eat for dinner to discussing political controversies; from choosing a favourite song to negotiating a pay rise.</p>	<p>Sentence (or idea unit) level, discourse level organisation, pragmatic and sociolinguistic awareness, interactive language.</p> <p>Discussions can elicit a wide range of language functions, although by their nature they are rather unpredictable (a problem from the point of view of reliability).</p> <p>As with other speaking tasks, support can be provided by varying the topics or by providing guiding questions and visual support.</p>	<p>Well-chosen discussion topics that interest the learners are perhaps the most successful way to foster spontaneous natural communication in classroom settings. This is also their major weakness: not everyone is equally interested, willing or able to discuss certain topics.</p> <p>Discussion can be carried out at a whole class level, but learners usually perform better and will have more opportunity to speak when working in groups of three.</p> <p>In the classroom, a choice of topics can be given so that everyone has the chance to demonstrate their abilities, but giving choices reduces comparability and threatens reliability. In tests, either no choice of topic is offered, or the choices are very limited.</p>
Role Play	<p>Role playing is an option that may be brought into any of the above task types.</p> <p>Assesseees may, for example, present a monologue in the role of a salesperson, take part in a discussion in the role of an opponent of a new building project or play the part of an employer in a job interview.</p>	<p>A key advantage of role playing is that it can reflect a variety of social roles and speech events that are not easily captured any other way. These include, for example, interactions between doctors and patients, tour guides and tourists, air traffic controllers and pilots, sales assistant and customers. This makes role play particularly appropriate in assessments of languages for specific purposes.</p> <p>Role play can help to extend the kinds of interaction elicited beyond those that would naturally occur between learners and an assessor.</p>	<p>Most people are familiar with a variety of roles including being a customer in a hairdresser's or a patient in a doctor's office. The role of the hairdresser or doctor will be less familiar and therefore more challenging for most learners.</p> <p>It may be more difficult (from a personal rather than a linguistic point of view), to take on the character of another person. For example, some role plays call for the assessee to act angrily or to express a certain opinion. In these cases, construct-irrelevant factors – like acting ability – may come into play and affect the scores.</p>