This is a repository copy of *Assessing depth of word knowledge of beginner learners of French, German, and Spanish aged 11-14 in England*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/184516/

Version: Submitted Version

Book Section:

Finlayson, Natalie and Marsden, Emma orcid.org/0000-0003-4086-5765 (2023) Assessing depth of word knowledge of beginner learners of French, German, and Spanish aged 11-14 in England. In: Chong, Sin Wang and Reinders, Hayo, (eds.) Innovation in Learning-Oriented Language Assessment. Palgrave Macmillan , pp. 137-161.

https://doi.org/10.1007/978-3-031-18950-0

To cite this chapter:

Assessing depth of word knowledge of beginner learners of French, German, and Spanish aged 11-14 in England

Natalie Finlayson and Emma Marsden

National Centre for Excellence for Language Pedagogy[1], Department of Education, University of York

Abstract

This chapter focuses on the development of vocabulary achievement tests for young, beginner learners of French, German and Spanish as foreign languages. Key challenges at this level include: (1) identifying test items that correspond to relevant aspects of word knowledge; (2) adapting the design of these items to the needs of school pupils; (3) ensuring parity across learners and languages; (4) designing flexible tests that can be completed both in-class and online. We describe our test purpose and relevant rationales, and present examples of items used in England's National Centre for Excellence for Language Pedagogy (NCELP) tests, which are currently being piloted in some of NCELP's 45 networked schools and beyond.

1. Introduction

The National Centre for Excellence for Language Pedagogy (NCELP) works with researchers, teacher educators, and specialist practitioners to improve curriculum design and pedagogy in schools in England. It is funded by the Department for Education to provide research-informed continuing professional development and classroom resources to support teachers' understanding and delivery of recommendations made in the Modern Foreign Languages (MFL) Pedagogy Review of teaching practice in Key Stages (KS) 3 and 4[2] (Teaching Schools Council [TSC], 2016). Recommendations for the teaching and sequencing of vocabulary at this level are supported by research that emphasises the importance of (i) teaching individual, high-frequency words over purely topic-led and heavily formulaic language, (ii) developing depth of word knowledge, and (iii) a focus on structured recall. Recommendations for assessment are underpinned by evidence in favour of a more tightly-focussed approach to testing that measures knowledge of specific words in partial isolation from other language features. This provides teachers with information about whether students' knowledge is ready for use in more communicatively demanding or time-pressured contexts, and in inferring the meaning of unknown words.

Here, we first provide an overview of the KS3 context and discuss associated challenges for vocabulary teaching and testing. We reflect on what a 'learner-centred approach' might mean for KS3, and identify an appropriate theoretical framework for vocabulary test design. We present examples of how these situational and theoretical considerations are operationalised in NCELP vocabulary tests, and look ahead to further work. The term 'English' (rather than 'L1') is used throughout as not all students in schools in England have English as a first language.

2. The KS3 Context: Learners of French, German and Spanish aged 11-14 in the UK

MFL teaching is compulsory for pupils aged 7-11 in English primary schools, but provision is not yet well established. Thus, it is difficult to know what language knowledge students bring to KS3 (TSC, 2016; British Council, 2021). Although a small proportion of students may already have reliable knowledge and grammatical terminology, KS3 is crucial for developing intrinsic motivation, language knowledge, and good learning habits for the majority. Inter- and (for different languages) intra-school variation in the number of curriculum hours devoted to MFL further complicates this disparity; on average, the range is one to four lessons of 40-60 minutes per week. Thus, NCELP had to make a pragmatic decision to create resources - including tests - based on a notional average of two 50-minute sessions per week, with the expectation that these could be extended or condensed depending on individual contexts.

Currently (until the GCSE exams in 2026), there are no fixed wordlists for KS3 and 4, and tests of isolated vocabulary items are not part of GCSE examinations. Awarding organisations may choose to produce wordlists if they wish, but are not required to test the words they contain. Perhaps because of this, systematically designed vocabulary tests are not provided in widely used textbooks (e.g., *Studio, Stimmt!, Viva* (Pearson); *Zoom, Équipe Nouvelle*

---

[2] Key Stages 3 and 4 are stages of secondary education in England and Wales. KS3 comprises Years 7-9 (ages 11-14, at which stage study of an MFL is obligatory) and KS4 Years 10-11 (ages 14-16, when about half of students choose to study an MFL, approximately 250,000 pupils in England). At the end of KS4, students sit a high stakes external examination, The General Certificate of Secondary Education (GCSE).

(Oxford University Press)). Many teachers provide topic-driven lists of words and phrases (e.g. hobbies, family, food and drink), and check knowledge of these in short, low-stakes class tests such as written translation of individual words from L2 into English, and general tests of comprehension and production (TSC, 2016). It is not common practice for these learning activities to be followed by isolated and closely-linked vocabulary assessments for school reporting; instead, topic-based proficiency tests are used. The tests we describe in this chapter are innovative for current practice, and demonstrate the kinds of assessment that might be used in a KS3 curriculum that is more language driven than hitherto (DfE, 2022).

We now consider the practical challenges that restrictions on curriculum time and mixed prior knowledge pose to the development of KS3 vocabulary tests for widespread use and adaptation by teachers.

2.1 Curriculum time

Cross-school variation in time available for MFL instruction has implications for (i) the amount and type of material that can be taught and assessed, (ii) the practicalities of test administration, and (iii) the ease and speed with which constructive feedback can be delivered.

Milton (2006) found that 16-year-old pupils had receptive knowledge of 852 words after 5 years of study, averaging at an acquisition rate of about 4 words per lesson. Schmitt (2000) considered the introduction of around 10 words per 1-hour lesson to be a reasonable target. As a broad indication for curriculum planning, a rate of 5-6 words per lesson allows for the teaching, practising and sufficient revisiting of around 1,090 words over the course of KS3[3]. This figure might seem low at first glance, but is underpinned by the notion that acquisition of 1,000 well-chosen words could equip leaners with a reliable core vocabulary (see 3.1), and by the fact that after five years of learning, approximately 1,700 words seem learnable, in line with the revised GCSE subject content (DfE, 2022).

At KS3, language tests for school reporting are often delivered termly (two or three times per year, though this varies between schools), with one or two lessons at each testing period dedicated to summative assessments. As vocabulary is only one component of language assessment (alongside, for example, grammar, phonics, and use in context), it is plausible that the total time available for vocabulary testing across a school year could be as little as 50 minutes. Approaches to KS3 vocabulary assessment, then, must be quick to administer, complete, and mark, and provide sufficiently detailed information about students' knowledge of a target set of words. One step towards achieving this is to identify test items that can be completed on computers and marked computationally. Online test development became a priority for NCELP during the pandemic, and some teachers continue to use these tests, perhaps because they can save time and support inclusive learning. However, the type of test that can be administered online affects various aspects of validity, and is influenced by the theory of assessment upon which it is drawn (see section 3).

2.2 Parity across tests and cohorts

Other issues relating to parity are those posed by inter-student variables that can advantage certain groups of learners over others. These include extracurricular language learning

---

[3] Based on a 36-week school year.

experience, heritage knowledge of the L2, and the range of cognitive abilities in any given cohort. Another key issue is interlingual structural variation. The phonological, orthographical, morphological, syntactic and semantic systems of French, German and Spanish, and their relations with English, each present unique challenges which must be reflected in test design. We have aimed to achieve equity by focussing on a carefully-selected set of words (see 3.1), testing only these words, and adapting established test types with rubrics tailored to the needs of young beginners. We have also emphasised to teachers that our tests are examples (prototypes) which they will need to adapt to their own contexts.

2.3 Learner-centred assessment for KS3

It seems appropriate at this point to offer a definition of 'learner-centred assessment' that acknowledges the context of this young and diverse group of students. For many students, KS3 represents the first steps of the journey towards becoming an autonomous language learner, and we aim to support this by empowering them with a core vocabulary that is useful in a range of contexts. We do not consider these inexperienced students to yet know enough about their language learning needs to involve them in assessment design directly, but we consider our approach to be learner-centred for several reasons:

(i)   it provides clearly defined lexical goals and space in the curriculum to practise the lexicon,
(ii)   it supports the teaching and testing of words that are relevant to the general needs of the target group,
(iii)   it aims to promote intrinsic motivation by teaching learners to manipulate a core vocabulary,
(iv)   it provides learners with content that facilitates 'ownership' of a personal lexicon that is important for them, but is unlikely to be needed by other members of the class (e.g. pets, parents' jobs),
(v)   it supports a formative approach to assessment with plentiful opportunities for detailed feedback and preparation for a 'new' style of tests.

In sum, we established a need for an approach to KS3 vocabulary assessment that supports learner development and intrinsic motivation by:

- testing a manageable set of carefully-chosen words
- testing word knowledge in a way that allows meaningful feedback to be given
- designing tests that support parity and are quick to administer and mark.

These parameters determined the theories of assessment relevant for operationalising these objectives, which are discussed now.

3. Assessment theory

Vast bodies of testing research look at the vocabulary knowledge of monolingual or multilingual children, and adults with English as a second language. However, work on designing vocabulary tests for *children* learning *languages other than English* (LOTE) is rare (though see Meara & Milton, 2003 for a yes/no lexical decision test in various languages; and tests used by Peters et al., 2019). Until further evidence is gathered, creators of LOTE tests for young or early-stage learners must adapt tests designed for other purposes, as we have

had to do. A full review of vocabulary assessment literature is beyond the scope of this chapter, and we focus here on three core considerations: (i) the number and nature of words to test; (ii) the type of test; (iii) the question style.

3.1 Deciding what to test

First, we explore key matters of lexical selection: frequency, aspects of word knowledge, and single words versus formulaic sequences.

3.1.1 High- and low-frequency words

There is considerable agreement over the degree to which the acquisition of a relatively small number of word families can facilitate general understanding of an L2. Research on text coverage in English shows that gains in comprehension (or 'coverage') of general texts plateau significantly once learners have mastered the 2,000 or so most frequent words (Milton, 2006; Nation, 2013). It seems logical, then, that a core wordlist for beginners would contain a large proportion of high-frequency words, but this is not the case in English schools. A heavily topic-driven approach to vocabulary selection and subjective decisions about which words are of interest to students has led to the teaching of rare vocabulary that is only useful in specific contexts (Häcker, 2008). This is probably at the expense of giving students robust and reliable access to commonly occurring words that (i) are applicable in a range of situations, (ii) reflect core features of the target language, and (iii) give insights into the target culture (Kilgarriff et al., 2014). A frequency-informed approach to vocabulary selection addresses these issues, and provides an objective index for working towards cross-language parity when deciding what the most useful words might be.

This observation has two important caveats. First, the 2,000 most frequent words in general language are not constant; beyond the top 500 words or so, wordlists start to reflect the content of the corpora from which they are derived. For this reason, we have selected vocabulary from frequency lists comprising words with high frequency in several general corpora (as per Brezina & Gablasova, 2015). The Routledge frequency lists of French (Lonsdale & Le Bras, 2009), German (Tschirner & Moehring, 2019), and Spanish (Davies & Davies, 2018) provide the frequency order of words in our Schemes of Work (NCELP, 2021a), and we have used the Web as Corpus (WaC) and TenTen Corpus Family as sources for preliminary corpus-based evaluations of the words they contain. Second, not all high frequency words are relevant to the needs of schoolchildren. Thus, we estimate that 85% of words on a KS3 list could usefully be taken from core general frequency lists, and supplemented by a smaller set of carefully-chosen lower-frequency words. As per DfE (2022), awarding organisations will compile wordlists for the GCSE, and we have recommended that they use teacher perspectives (Dang et al., 2020) and lexical coverage measures (of texts used later in the curriculum e.g., exams) as additional measures of evaluation.

We acknowledge that no list, however long, can represent all individual interests of KS3 students, and as part of a learner-centred approach we encourage the development of dictionary skills to support a personal lexicon (see NCELP, 2021b for an example). For reasons of curriculum time and parity, however, we do not include these additional words in tests.

3.1.2 Aspects of word knowledge

In 2.1, we recognised a need for an approach to KS3 vocabulary testing which supports the systematic development of knowledge of a relatively small set of words. Considering word knowledge as the mastery of Nation's (2013: 60) three core components of word knowledge - form, meaning, and use - seems a useful approach to doing this. These components can be further broken down into nine aspects of word knowledge ordered broadly by complexity, representing a structure for subcomponents of knowledge that develop as words are consolidated. This could help practitioners identify the most important teaching and testing goals for their target group.

Knowledge learned in the early stages is likely to comprise *written and spoken forms* and *form and meaning connections.* Revisits throughout KS3 provide opportunities to deepen understanding by developing knowledge of *word parts*, *concepts and referents*, *associations* and some *constraints on use.* Structuring vocabulary teaching and testing around these aspects of word knowledge facilitates systematic revisiting in the curriculum, the value of which is undisputed (Nation, 2013; Schmitt, 2000).

3.1.3 Individual words and formulaic sequences

An emphasis on individual words - with a complementary reduction on formulaic sequences (chunks) - aims to help learners manipulate words and grammar, and understand and use component parts of sequences in a range of contexts. This supports the goal of language parity in three ways.

First, a focus on single words can help to offset the complexity caused by the grammatical or semantic contexts experienced during incidental learning. In such approaches, learners' individual analytic ability may determine whether they can 'pick out' and interpret words from the input. For example, *je m'* (I…myself) may be mistakenly parsed as "I am" if *je m'appelle* (I'm called; literally, I call myself) is not broken down for learners. Second, it supports the careful selection of English translations for testing, which is especially important for high-frequency words which are often polysemous (e.g. *historia* (history, story)). Third, it allows highly irregular forms (e.g. *ich bin* (I am)) to be treated as individual lexical items, an approach informed by the notion that these bear little resemblance to the word family headword, and are likely stored as separate words in the mental lexicon (see Meunier & Marslen-Wilson, 2004 for related arguments). Finally, it allows consideration to be given to part of speech proportions, mitigating a consequence of a topic-driven approach which can skew proportions towards nouns (Häcker, 2008; Marsden & David, 2008).

3.2 Test design considerations

At the heart of test validity lies a clear definition of the test purpose and the construct that it will measure (Bachman & Palmer, 1996). We consider this now in a review of test types, measures, and question styles pertinent for our learner group.

3.2.1 Types of test

In the context of vocabulary testing, *placement* tests estimate the number of words a learner knows, *achievement* tests assess knowledge of a set of words taught in a course, and

proficiency tests measure overall language level (Nation, 2013). For achievement tests, the current focus - alignment of vocabulary taught and assessed - is crucial for validity; that is, responses elicited from learners should represent knowledge of the words being tested, and not be compromised by the presence of unknown words (Schmitt, 2000). Given the limited vocabularies of KS3 students, and the guiding principles of parity, concision, and feedback opportunities, a syllabus-based achievement test seems most appropriate for our purposes. Some test researchers (e.g., Read, 2000) extend the definition of achievement testing to include assessing mastery of skills, such as lexical inferencing. NCELP do not test broader applications of vocabulary in the very early stages, and examples of this kind of test are not included in this chapter. However, inference tests are introduced towards the end of KS3, when students' core vocabularies are wider.

3.2.2 Measures of vocabulary knowledge

Vocabulary *breadth* (size) tests indicate the overall number of words known by a learner, while *depth* tests assess knowledge about a set of target items. Size tests are less useful at KS3 because of the relatively small number of words students learn at this level. Schmitt (2000) considers depth tests a more valid assessment format, as they contain a mix of question styles and address a fuller, more precise range of lexical aspects. Breadth tests traditionally comprise multiple choice, gap fill, yes/no, or single-word translation exercises that are either receptive or productive and test only one aspect of knowledge. Depth tests assess both receptive and productive uses of words in a variety of ways, allowing demonstration of partial word knowledge. This complements the goals of KS3 in that it is motivating for the learner, comes closer to giving an indication of the true extent of lexical knowledge, and provides opportunities for feedback on a wider range of knowledge types.

3.2.3 Question styles and level of difficulty

Recognition items (involving the use of choices) have been shown to be easier than recall items (requiring the provision of form or meaning from memory), and receptive tests more straightforward than productive ones (Nation, 2013). It might be expected, then, that lower-level tests comprise a proportionately higher number of recognition items. However, research points to the benefits of desirable difficulty (Suzuki, Nakata, & DeKeyser, 2019) and practice in recall (Kemp & MacDonald, 2021). As formative tests are part of the study process, promoting recall practice both prior to and during assessment is likely to benefit learning.

Multiple-choice items have been found to be easier than items with response formats such as gap fill, free production, and translation (Nation, 2013). They are quick to mark and transferrable to online testing, suggesting their suitability for our context. Carefully-chosen distractors can be used to control the level of difficulty of multiple-choice items. The closer the distracters are in form and meaning, the more difficult the item becomes (Kemp & MacDonald, 2021). Self-report formats (yes/no tests) are not recommended because of the restricted information they provide about word knowledge (Read, 2000).

The use of English in question types and rubrics is appropriate in the early stages, becoming less appropriate as knowledge advances (Read, 2000). This consideration affects decisions over whether and how to present target items in context. Read recommends the use of context for (i) signalling the intended meaning of a target word, and (ii) eliciting a target item in more

challenging, monolingual questions. However, he notes the value of testing items in isolation to align with memorisation techniques used by students in systematic vocabulary learning, and for determining the extent to which students can demonstrate knowledge of words without contextual support.

3.3 Representative sampling

To create representative samples from wordlists, decisions must be made about proportions of the following: revisited words and 'new' words introduced since the previous test; word classes; listening, reading, writing and speaking activities; recognition and recall activities. The proportions are likely to change with progress through KS3. For example, NCELP tests include a roughly equal proportion of new: revisited words in the early stages, but include fewer new items in Year 9 when the body of revisited vocabulary is larger.

Care must be taken to ensure students do not need to draw on grammatical knowledge to answer questions correctly, even when the items test isolated words. This is achieved by testing headwords as they appear on the wordlist (infinitive forms of verbs, masculine forms of nouns and adjectives, highly irregular forms treated as individual words).

3.4 Statement of test purpose

Having reviewed assessment literature relevant to the KS3 context, we are now in a position to define the test purpose. Our aim is to create syllabus-based, depth of knowledge achievement tests for French, German and Spanish that assess relevant aspects of knowledge of a sample of items on the NCELP KS3 wordlist (2021c).
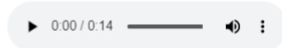
4. Example items from NCELP tests

Nation (2013: 538) summarises what is assessed when we test each aspect of word knowledge. We now present examples of NCELP test items that align with his framework, and show how the KS3 context and associated theoretical considerations inform their design. Examples represent a range of modes, modalities, and question types, and illustrate progression across KS3. Teachers are provided with detailed mark schemes for paper tests. The self-marking online versions of reading and listening items are available via the survey-hosting platform *Qualtrics* (see Figures 1 and 2). Productive items, which require marker sensitivity to (and tolerance for) spelling, word order, and pronunciation, are not available in electronic form.

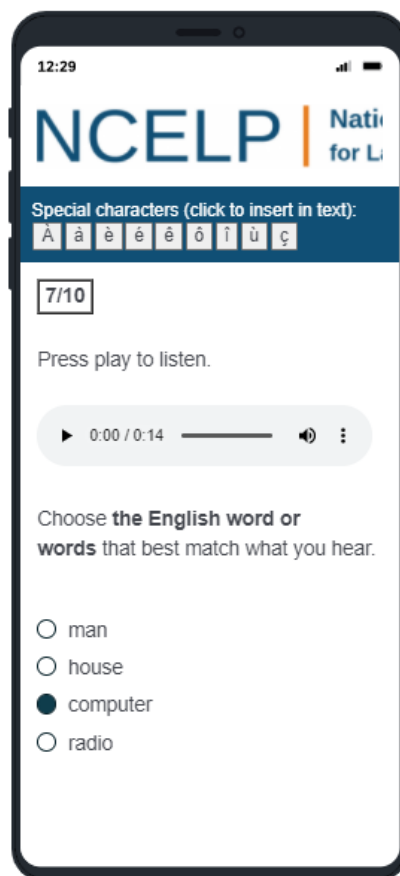Figure 1: Example of an NCELP test item on a web interface

Figure 2: Example of an NCELP test item on a mobile phone

With each example, we include a suggested rubric and audio transcripts where applicable. Based on teacher experience and preliminary piloting, we recommend an audio speed that is no faster than moderate speed. Captions comprise a description of the item type, the name of

the school term in which the test is taken (according to the NCELP Schemes of Work), and references to any established test on which it is based.

4.1 Testing aspects of form

Receptive tests of spoken form involve hearing an L2 word and choosing the correct English translation from distractors with varying degrees of attractiveness (Figures 3 and 4). In Figure 3, distractor items differ in both length and spelling. In Figure 4, the three translations beginning with the letter C are likely to appeal, with the two single-syllable words probably most attractive within those. We strived to increase the proportion of these more challenging items as students progress.

INSTRUCTION FOR STUDENTS

You will hear ten French words.

Put a **cross (x) under the English word or words** that best match what you hear.

You will hear each French word **twice**. Choose **one** correct answer only.

TRANSCRIPT

[Each word plays for the first time, 5 second pause, then plays for the second time].

1. l'ordinateur

6.	man	house	computer	radio
	☐	☐	☐	☐

Figure 3: Receptive test of form and meaning recognition (French Year 7, Term 2)

INSTRUCTION TO STUDENTS

You will hear ten Spanish words.

Put a **cross (x) under the English word or words** that best matches what you hear.

You will hear each Spanish word **twice**. Choose **one** correct answer only.

TRANSCRIPT

[Each word plays for the first time. 5 second pause, then plays for the second time.]

1. la moneda

1.	camera	phrase, sentence	coin	class
	☐	☐	☐	☐

Figure 4: Receptive test of form and meaning recognition (Spanish Year 7, Term 2)

Our approach to testing spoken form (Figure 5) reflects the importance of oral recall in the learning process (see 3.3). Recall is cued using English translations as students are used to seeing them on their wordlist, and students are told how many L2 words they need to say. Full marks are awarded for a response with no errors, half a mark for an item with one or two errors, and no marks for an item with three or more errors. An error is defined as an incorrect, omitted or indiscernible article, or an incorrect, omitted or added sound. A foreign accent is tolerated as long as it does not impede understanding.

INSTRUCTION TO STUDENTS

Say the **French** for the words below. **Remember** to say the word for '**the**' if needed!

1. the thing   (**two** French words)

2. to do/make, doing/making   (**one** French word)

3. amusing, entertaining (f.)   (**one** French word)

4. well   (**one** French word)

5. sad   (**one** French word)

Figure 5: Productive test of spoken form recall (French Year 7, Term 2)

The teaching of word parts (derivational morphology) for reading only is introduced in Year 8 German and Spanish, and Year 9 French[4]. The example in Figure 6 requires students to break an unknown word consisting of a known base and known affixes into parts. Figure 7 shows a more advanced form of this question style in which students must recognise which affixed form of a known L2 base completes a gapped sentence. Production of affixed forms falls outside the NCELP KS3 Scheme of Work, and is not tested.

INSTRUCTION TO STUDENTS

Read each French word.

Write the parts of the French word which contain the English meanings in **bold**.

1. quatrièmement

a) Which part of this word means '**-ly**' ? _____

b) Which part of this word means '**-th**' ? _____

Figure 6: Receptive test of word part recognition (French Year 9, Term 3)

---

[4]The later introduction in French recognises the additional learning burden of the larger number of phonics items introduced in Years 7-8, in keeping with NCELP's commitment to creating parity by using the language itself as the key driver for the curriculum.

Put a **cross (x)** next to the **one word** that best completes the sentence.

1.	Nur eine Antwort ist _____.	a) möglich	☐
		b) möglicherweise	☐
		c) Möglichkeit	☐

Figure 7: Receptive test of word part recognition (German Year 8, Term 3)

NCELP vocabulary assessments do not include tests of written form, as reading aloud and dictation skills are tested separately as part of phonics assessment.

4.2 Testing aspects of meaning

Of the three core components of word knowledge, meaning is arguably the most straightforward to teach at beginner level. This is especially true of the *form and meaning* aspect, which is tested in a number of ways across NCELP KS3.

Figure 8 shows a reading test in which students translate an L2 word into English. Context is provided to clarify the grammatical function of the target word, and care is taken to ensure that co-text does not give lexical hints.

INSTRUCTION TO STUDENTS

Translate the **underlined Spanish words** to complete the English sentence.

1. La **naturaleza** es hermosa.   _____ is beautiful.   (write **one** word)

2. Queremos la **comida.**   We want the _____.   (write **one** word)

Figure 8: Receptive test of form and meaning recall (Spanish Year 7, Term 2)

Figure 9 shows progression of form and meaning knowledge in a monolingual test of synonyms and antonyms. To ensure grammatical knowledge is not tested, all distractors are in the same word class and form as the target word. This monolingual format allows more words to be tested, but teachers must take care when giving feedback to ensure that students are clear where any errors lie (i.e., with understanding the target word, or the distractors).

INSTRUCTION TO STUDENTS

Put a cross (x) next to the word with the **most similar** meaning to the word in bold.

1. Es **alegre**.	a) aburrido	☐
	b) serio	☐
	c) feliz	☐
	d) importante	☐

Put a cross (x) next to the word which has the **opposite** meaning to the word in bold.

1. **Siempre** hace los deberes.	a) sólo	☐
	b) a veces	☐
	c) nunca	☐
	d) normalmente	☐

Figure 9: Receptive test of form and meaning recognition (Spanish Year 7, Term 3)

The exercises in Figures 10 and 11 test understanding of the collective meaning of L2 words in short definitions. Again, consideration is given to the selection of distractors. In Figure 10, answer options have varying degrees of relevance to *fotos* (photos), giving students a chance of success even if *foto* is the only part of the sentence they recognise. In Figure 11, there is an appealing distractor for each definition (two family members, two units of time, two items of clothing). Note the use of the word 'best' in the rubric (here and elsewhere) to allow for several answers potentially fitting, but only one being the most appropriate.

INSTRUCTION TO STUDENTS

You will hear a **short definition** in Spanish. Put a **cross (x) under the English word** that **best matches** the Spanish definition that you hear.

You will hear each Spanish definition **twice**.

TRANSCRIPT

[Say the number **in English** and leave a **1.5 second** pause before reading the Spanish definition. Leave a **5 second pause** and read the Spanish definition for the second time. Leave a **5 second pause** before moving on to the next Spanish definition.]

    1. Una cosa para sacar fotos.

1.	magazine	scene	camera	box
	☐	☐	☐	☐

Figure 10: Receptive test of form and meaning recognition (Spanish Year 8, Term 3)

INSTRUCTION TO STUDENTS

Write the letter (**a-f**) of the word **on the right** that **best matches** the definition on the left.

	Definition	Words
1.	___ Bruder und Schwester ___ sieben Tage ___ Kleidung für den Kopf	a. die Woche b. die Eltern c. das Monat d. der Schuh e. die Geschwister f. der Hut

Figure 11: Receptive test of form and meaning recognition (German Year 8, Term 2) (based on Read, 1995)

Written tests of form and meaning involve translation from English into the L2. In the early stages, this kind of item is contextualised in both languages (Figure 12). A more complex version provides context in the L2 only (Figure 13) to clarify that students must translate *school* as an adjective complement (and not as a noun). Students are not required to agree adjectives with noun genders; full marks are awarded for a correct masculine, feminine or plural adjectival form. In line with the notion of interlanguage scoring, tolerance is applied for missing or incorrect accents in Years 7-8, In Year 9, we suggest deducting half-marks for such errors.

INSTRUCTION FOR STUDENTS

Translate the **underlined words** to complete the Spanish sentence.

1. **The sentence** is correct.   ___ _____ es correcta (write **two** words)

Figure 12: Productive test of form and meaning recall (Spanish Year 7, Term 2) (Adapted from Nurweni & Read, 1999)

INSTRUCTION FOR STUDENTS

Translate the **English words in brackets** to complete the French sentence.

13. L'équipe _____ joue aujourd'hui. (**school** (adj.))   (write **one** word)

Figure 13: Productive test of form and meaning recall (French Year 9, Term 1)

In Figures 14 and 15, form and meaning are tested at greater depth than in their arguably less complex counterparts (Figures 4 and 13 respectively). In these tests of concepts and referents,

students must demonstrate knowledge of multiple meanings by recognising and producing more than one equivalent of an L2 or English target item.

INSTRUCTION TO STUDENTS

You will hear ten Spanish words.

Put a **cross (x) under the English word or words** that best match what you hear.

Some **have only one correct answer**. Some have **two correct answers**.

You will hear each word in Spanish **twice**.

TRANSCRIPT

[Each word plays for the first time, 5 second pause, then plays for the second time.]

5. listo

5	strange	intelligent	ready	calm
	☐	☐	☐	☐

Figure 14: Receptive test of concept and referent recognition (Spanish Year 8, Term 2)

INSTRUCTION TO STUDENTS

Write **two Spanish words** for each of the following English words:

1. to send   1. _____, 2. _____
2. woman   1. _____, 2. _____
3. beautiful   1. _____, 2. _____
4. to take   1. _____, 2. _____

Figure 15: Productive test of concept and referent recall (Spanish Year 9, Term 3)

Figure 16 shows an association test at the level of hierarchical organisation. Students must assign an L2 target item to an English category headword. Figure 17 shows an adapted version of Read's (2000) Word Associates Format, in which the somewhat complex format of the original is split so that students think about meaning and collocation separately.

INSTRUCTION TO STUDENTS

You will hear 10 French words. **Put a cross (x)** under the **type of word** you hear.

You will hear each French word **twice**.

TRANSCRIPT

[Say the number **in English** and leave a **1.5 second pause** before reading the French word. Leave a **5 second pause** and read the French word for the second time. Leave a **5 second pause** before moving on to the next word].

3. la sœur

This word is a good example of …			
3.	a type of food ☐	an object ☐	a family member ☐	a school subject ☐

Figure 16: Receptive test of association recognition (French Year 7, Term 2)

INSTRUCTION TO STUDENTS

Put a cross (x) next to the word with the **closest related meaning** to the word in bold.

1. das Haus	a) im Chor	☐
	b) tragen	☐
	c) die Flasche	☐
	d) wohnen	☐

4. das Paar	a) kein	☐
	b) wiederholen	☐
	c) das Eis	☐
	d) zwei	☐

Put a cross (x) next to the word which could appear **beside the word in bold** in a sentence.

1. kochen	a) benutzen	☐
	b) das Konzert	☐
	c) das Fenster	☐
	d) Gemüse	☐

3. sagen	a) ihr	☐
	b) das Wort	☐
	c) die Hose	☐
	d) suchen	☐

Figure 17: Receptive test of association (German Year 7, Term 3) (Adapted from Read, 2000)

4.3 Testing aspects of use

The most complex of the word knowledge elements are the aspects of use. Students may acquire some collocational knowledge through learning important multi-word expressions, but we are not confident that this knowledge is promoted by KS3 NCELP resources, or generalisable and testable at this level. As KS3 progresses, we do expect students to have developed some awareness of certain examples of register, and perhaps of regional variation. Figure 18 illustrates how knowledge of register may be tested productively. This item could be

adapted to test, for example, knowledge of Spanish words in the Castilian and Latin American varieties.

INSTRUCTION TO STUDENTS

Say the following informal words or phrases using more **formal language.**

1. la maman   2. tu vas

Say the following formal words or phrases using more **informal language.**

3. bonjour   4. vous êtes   5. s'il vous plaît

Figure 18: Receptive test of constraints recall (French Year 8, Term 1)

5. Conclusions and next steps

We have identified a need for KS3 vocabulary assessments that support development and intrinsic motivation through the efficient testing of aspects of word knowledge in ways that align with course objectives. With reference to relevant assessment theory, we have shown how these aims are operationalised in NCELP achievement tests for KS3 French, German, and Spanish.

We hope to analyse response data to ascertain test reliability, and preliminary analyses show medium to high reliability indices (Cronbach's alpha). However, as we hoped and intended, use of NCELP resources inevitably varies to suit context and teachers' and learners' requirements and so we cannot be sure about the extent to which test data reflect the purpose of the assessments as *syllabus-based achievement* tests. Long-term plans include analyses of the relationships between teachers' experiences of professional development, the extent and nature of how they subsequently use the resources with their students, and ultimately the students' outcomes as measured by the kinds of tests described in this chapter.

The development of integrated skills tests (e.g., listening and reading comprehension, translation, picture description) and KS4 assessments is ongoing.

6. Practice brief

Adaptable achievement tests for KS3 French, German and Spanish are freely available on the NCELP resource portal: https://resources.ncelp.org/. In recognition of the fact that vocabulary testing is not currently part of every teacher's toolkit, the assessment collections also contain example lessons comprising activities similar to the test items. These can be integrated into course materials to familiarise students with the question styles. The NCELP Teacher Research Group professional development session on assessments (2021d) provides guidance for teachers on adapting our tests for different contexts.

Informal feedback from teachers on the testing materials referred to in this chapter suggests that it is useful to carry out the whole class parts of the assessment (listening, reading, writing) in one lesson, and the speaking sections in another. Online versions of the tests and automated marking seem to work well, but rely on a stable internet connection. For these, it

may be useful to carry out the parts of the test which require equipment (headphones and microphone) in one session, and the reading and writing parts in another.

References

Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice.* Oxford University Press.

Brezina, V., & Gablasova, D. (2015). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics, 36*(1), 1–22.

British Council. (2021). *Language Trends 2021.* https://www.britishcouncil.org/sites/default/files/language_trends_2021_report.pdf

Dang, T. N. Y., Webb, S., & Coxhead, A. (2020). Evaluating lists of high-frequency words: teachers' and learners' perspectives. *Language Teaching Research, 24*(1), 1-25.

Department for Education. (2022). *French, German and Spanish: GCSE Subject Content (January 2022).* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1046043/GCSE_French_German_Spanish_subject_content.pdf

Häcker, M. (2008). Eleven pets and 20 ways to express one's opinion: the vocabulary learners of German acquire at English secondary schools. *Language Learning Journal, 36*(2), 215-226.

Kemp, L.S. & McDonald, J.L. (2021). Second language vocabulary acquisition: The effects of semantic relatedness, form similarity, and translation direction. *Language Learning, 71*(3), 716-756.

Kilgarriff, A., Charalabopoulou, F., Gavrilidou, M. et al. (2014). Corpus-based vocabulary lists for language learners for nine languages. *Language Resources and Evaluation 48*(1), 121–163.

Marsden, E. & David, A. (2008). Vocabulary use during conversation: A cross-sectional study of development from Year 9 to Year 13 among learners of Spanish and French. *Language Learning Journal 36* (2), 181-198

Meara, P. & Milton, J. (2003). *X_Lex: The Swansea Levels Test.* Express.

Meunier, F., & Marslen-Wilson, W. (2004). Regularity and irregularity in French verbal inflection. *Language and Cognitive Processes, 19* (1), 169–206.

Milton, J. (2006). Language Lite? Learning French vocabulary in school, *Journal of French Language Studies, 16*(2), 187-205

NCELP. (2021a). *Vocabulary lists: Rationales and uses*. National Centre for Excellence for Language Pedagogy: University of York, UK. https://resources.ncelp.org/concern/resources/t722h880z?locale=en.

NCELP. (2021b). *French SOW Y7 Term 3.2 Week 7 – full lesson – using word reference resources*. National Centre for Excellence for Language Pedagogy: University of York, UK. https://resources.ncelp.org/concern/resources/r494vk73v?locale=en

NCELP. (2021c). *NCELP Schemes of Work.* National Centre for Excellence for Language Pedagogy: University of York, UK. https://resources.ncelp.org/schemes-of-work

NCELP. (2021d). *TRG 2.2: Assessment Part 1.* National Centre for Excellence for Language Pedagogy: University of York, UK. https://resources.ncelp.org/collections/zc77sq31q?locale=en

Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes, 18*(2), 161-175.

Peters, E., Noreillie, A.-S., Heylen, K., Bulté, B. & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning, 69*(3): 747-782.

Read, J. (1995). Refining the word associates format as a measure of depth of vocabulary knowledge. *New Zealand Studies in Applied Linguistics, 1*(1), 1-17.

Read, J. (2000). *Assessing Vocabulary.* Cambridge University Press.

Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge University Press.

Suzuki, Y., Nakata, T. & Dekeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal, 103*(3), 713-720.

Teaching Schools Council. (2016). *Modern Foreign Languages Pedagogy Review.* https://pure.york.ac.uk/portal/files/54043904/MFL_Pedagogy_Review_Report_TSC_PUBLISHED_VERSION_Nov_2016_1_.pdf

Tschirner, E. & Moehring, J. (2019). *A frequency dictionary of German: Core vocabulary for learners.* Routledge.