

EAL ASSESSMENT FRAMEWORK FOR SCHOOLS: RESEARCH AND DEVELOPMENT

Underpinning the EAL Assessment Framework for Schools

Research and development – what underpins the framework?

The goal of the study

The Research and Development Team were commissioned by The Bell Foundation to provide scales of Proficiency descriptors to chart the progress of EAL students. Our work on the EAL Assessment Framework for Schools contributes to work done in a range of countries and over a number of years to characterise the learning paths of EAL students and thus to develop a shared understanding of how their progress can be best supported.

The EAL Assessment Framework for Schools has five bands, comparable to the five-stage scheme originally developed in Wales.

Saying ‘comparable to’ in the previous sentence begs the question of how scales of Proficiency can be compared. A key aim of this study has been to address this issue and to illustrate best practice in constructing and interpreting such scales.

We believe that there is benefit in researching existing published frameworks to establish a solid foundation for further development. To do this we have sampled descriptors from existing scales and ordered them using state-of-the-art techniques.

The principle of comparative judgement

Constructing the proficiency scales involves ordering a large number of performance descriptors, from the most difficult down to the easiest. This has been done using skilled human judgements, but of a very specific kind.

There are two kinds of judgement: *absolute*, e.g. ‘this student is at CEFR [Common European Framework of Reference for Languages] level A2’, and *relative*, e.g.: ‘this student is more competent than that student’. Humans are generally very good at making relative judgements, but generally much worse at making absolute judgements, and will often find themselves in disagreement.

Comparison is perhaps our most fundamental approach to making meaning: as Laming (2004) states, ‘there is no absolute judgement. All judgements are comparisons of one thing with another’.

Thus the psychometric procedure called Comparative Judgement (hereafter CJ) sets out to organise and standardise human judgement so as to play to its strengths - that is, by making relative rather than absolute decisions. While the principles of CJ have long been understood there seems to be a new interest in exploiting CJ to address assessment issues.

No More Marking

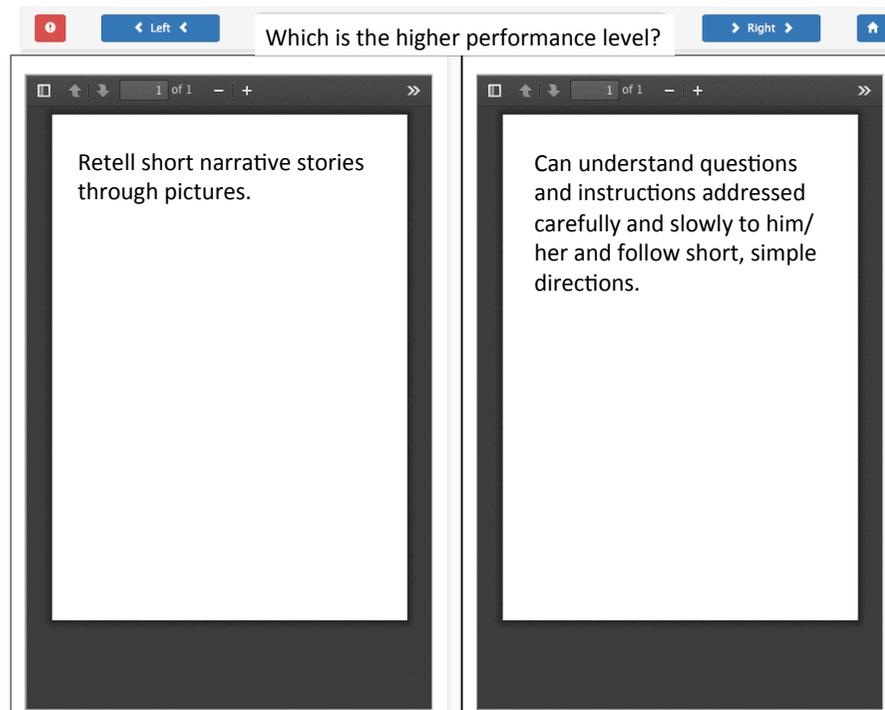
No More Marking (hereafter NMM) is the name of the website that hosted the data and the analysis for the descriptors contained in the EAL Assessment Framework for Schools. This site is promoting the use of comparative judgement to improve upon traditional approaches to marking used by assessment bodies in educational testing.

While the orientation of the NMM website is towards judging candidates in tests, the system can be used equally well for judging other entities - in our case, descriptors of EAL performance. The system is also flexible enough to enable working with several different comparative frameworks - in our study we were keen to treat the skills of Listening, Speaking, Reading and Writing separately, so that we could compare like with like.

The judging process

We invited teachers to make judgements on the NMM website. When teacher-judges log on to the system they see a series of paired descriptors, as illustrated in Figure 1. Their task is simple: to click against the descriptor that is the more difficult or advanced, that is, higher on the Proficiency scale. Then the next pair is presented. Each descriptor may appear several times, and each pairing adds to the reliability of the measure. Judges had a fixed allocation of judgements to make for each category (e.g. Secondary Reading), and most judges completed more than one category. In the notes provided to judges they were instructed as follows: ‘The strength of the approach is in the number of judgements made and the simplicity of the judgement. Understandably, judges begin slowly and carefully, and tend to worry when they cannot decide. But it is vitally important to work quickly and trust to first impressions ... Remember that each descriptor will be judged a number of times by a number of judges. Ten or fifteen seconds per judgement should be about right.’

Figure 1 the No More Marking judging screen



A large selection of assessment descriptors organised into a number of categories were uploaded to the NMM website. The descriptors were then presented to the participant teachers for their judgement

By and large judging worked smoothly, although there were some issues with judges understanding how to use their credentials to log into the system. In the early stages judges also pointed out errors in the categorisation of some descriptors, which were by degrees rectified.

The approach to analysis

The desired outcome was to construct for each language strand an ordering of descriptors on a calibrated measurement scale. How could the 11,755 judgements be processed to produce this outcome?

Item Response Theory (IRT) is the branch of statistics used in our approach. The essence of IRT is that it takes test scores and separates out the contribution to those scores of item *difficulties* and person *abilities*. These mutually-defining properties enable us to construct a scale, in a process called *additive conjoint measurement*. The scale measures invisible properties such as ability and difficulty in units, which is like a ruler measuring length in centimeters. The variant of IRT used by the No More Marking system is a specific one based on comparative judgements (the Bradley-Terry model). Essentially, the judgements in each paired comparison provide the basis for ordering the descriptors and estimating their precise difficulty (the column labelled "ability" in the table below).

The second column in Figure 2 is included here to remind us that the numerical ability values define a scale that can be transposed into meaningful levels. The text of the descriptors is shown truncated here.

Figure 2 An example of Bradley-Terry data

Ability	Working levels	Text
-3.0	Pre-A new to English	Can understand simple, straightforward language and ...
-3.0	Pre-A new to English	Can understand simple directions relating to how to get ...
-2.9	Pre-A new to English	Identify objects, figures, people from oral statements or ...
-2.9	Pre-A new to English	Listens to others for short periods of time.
-2.8	Pre-A new to English	Draw pictures in response to oral instructions.
-2.7	Pre-A new to English	Match oral language to classroom and everyday objects.
-2.7	Pre-A new to English	Can understand and use independently many of the ...
-2.6	Pre-A new to English	Can follow others to make choices in practical tasks and ...
-2.6	A new to English	Joins in willingly with routines.
-2.6	A new to English	Vocabulary is growing quickly and noticeably, fed by exper....
-2.5	A new to English	Point to stated pictures in context.

The guiding principles of EAL assessment

The development of the EAL Assessment Framework for Schools was guided by the following principles, drawn up by the Research and Development Team, Dr Michael Evans, Dr Neil Jones, Professor Constant Leung and Dr Yongcan Liu.

An assessment and evaluation framework for EAL

Collecting evidence to support EAL learners as they improve their proficiency in English should take account of individual and contextual factors. An EAL assessment and evaluation framework should support an evidence-based approach to assessing progress and to promoting learning, where evidence comes from multiple sources, including formal and informal assessment. Such an approach should have both quantitative and qualitative aspects: measurement of progress against a well-defined scale combined with a careful description of an individual pupil's knowledge, background experience and skills profile.

Such a framework must provide an orientation for learners and teachers as to each EAL learner's starting point, current level and potential development. Any EAL assessment framework should provide robust indicators of a learner's profile of English language knowledge and skills, strengths to build on and weaknesses to address.

The notion of progression in Proficiency is fundamental, whether we use terms such as 'standards', 'levels', 'phases' or 'stages'. The notion of progression allows a global scale against which all unique learner profiles can be evaluated, and which enables the development of a common understanding among practitioners. This is valuable for a range of teaching and administrative purposes, including curriculum development and teacher professional development.

An assessment framework for EAL should thus comprise both global measures and measures or evaluations of a range of aspects of language proficiency, relevant to the situation of EAL pupils. The framework should contain illustrations of pupil achievement through performance exemplars that characterise standards or levels in particular knowledge and skills. The complementary roles of classroom assessment and psychometric tests should be considered.

Key principles for an EAL assessment and evaluation framework

The following principles can be used to ensure that the impact of the assessment and evaluation framework is positive and effective.

Every EAL learner should achieve their full potential

EAL assessment should aim at ensuring that all learners achieve their full potential, with particular reference to English language development. The assessment and evaluation framework should be primarily formative in purpose. It should indicate what a pupil can do, inform curriculum provision, and enable diagnosis of needs and individualisation of learning, so as to promote potential development.

Every EAL learner has a unique profile

EAL learners are a heterogeneous group and each EAL learner has a unique profile that reflects the influence of the following factors:

Age

First language educational history

First language literacy

English language proficiency

English-medium schooling experience

Academic track record

Family and community circumstances.

EAL assessment should take full account of profile factors in the following ways:**Age**

Age should be taken into account because it strongly influences the kind of cognitive operations and/or communicative uses of language that it is reasonable to expect of a child at particular stages, other things being equal. A pupil first entering English-medium education at age five is likely to have very different social and curriculum communication needs to a pupil entering at age 14. EAL assessment should capture this in its performance descriptors.

First language educational history

EAL pupils arrive in the UK at different ages and with different life experiences. They may have studied to a level where they understand much of the content of a school subject, and may also be familiar with the notion that particular academic subjects adopt particular modes of discourse. A pupil's first language educational history can be seen as an indication of first language development (see First language literacy below) and prior curriculum attainment.

First language literacy

If EAL pupils are literate in their first language, this will likely facilitate acquiring literacy in English. EAL assessment should make every effort to gather information on pupils' first language literacy, which can be used to inform expectations of progress.

Academic track record

Like all pupils, EAL pupils will vary in their academic track record, in general and in relation to particular subjects. On starting school in the UK, their academic track record should be evaluated as far as possible taking account of any language difficulties. This information can help teachers make informed decisions on target setting and programmes of work.

Individual and family circumstances

EAL pupils' propensities for learning can be influenced by a variety of individual and family circumstances within the local community context. EAL assessment should take account of information on relevant issues such as giftedness, Free School Meal, and Special Educational Needs & Disability where appropriate; these issues can impact on pupils' academic performance and language learning. EAL assessment should also track pupils' dispositions towards learning over time as an important aspect of a pupil's profile.

Assessment and evaluation serve multiple purposes

Any EAL assessment should, at a minimum, include an initial profiling system and rating scales of performance. It must provide different forms of evidence, to serve the following functions:

Initial profiling

The above section listed the key features of the profile of an EAL learner. Most of these can be established initially through interview, with the learner and/or a parent (with interpreter support where appropriate); current state of English language proficiency might be determined by a suitably trained teacher, with reference to performance exemplars and descriptors, as well as through more formal assessment. Previous academic track records in particular subjects should be evaluated as far as possible taking account of language differences and/or difficulties.

Informing expectations

The initial profiling should reveal strengths as well as areas of need. It should provide the school with a clear expectation of what the pupil can achieve and what specific issues to address. Over time careful observation of the trajectories of EAL pupils with particular profiles, facilitated by an appropriate assessment framework, should enable better understanding of what can be achieved, and setting of appropriate targets.

Orientation

As already noted, any assessment framework must provide an orientation for both pupils and teachers as to each EAL pupil's starting point, current level and possible goals: where did we start, where are we now and where are we going? This is essential information for pupils if they are to take responsibility for their learning. It requires that there are clear criteria for assessment and decision-making, ensuring that pupils' use of English can be evaluated in terms of real-world communicative abilities.

Feedback into learning

A fundamental purpose of an assessment framework is to bring about better learning. The rating scales should provide indicators of the pathways of development. These pathways should inform feedback at various levels, to individual pupils and to teachers. This feedback should distinguish language issues from conceptual /content issues, and indicate where individuals might best focus attention. Evidence should inform teachers in terms of possible progression to assist pupils in becoming better at learning and more autonomous.

Feedback for management

An EAL assessment framework should provide information on pupil achievement and progress that can be aggregated for use by management and policy makers (as well as pupils and teachers): global indicators of learning progress and outcomes of EAL pupils. This information should be used for curriculum planning and development.

Key criteria for evaluating EAL assessment frameworks and rating scales

Progression:

- Does the framework clearly differentiate the stages of progression in acquiring English?
- Does the framework take account of differentiation in rate of progress in the different language skills, e.g. Speaking, Listening?
- Does the framework identify progression in terms of communication (i.e. using English to convey meaning in context)?
- Does the framework identify progression in terms of formal accuracy? Is account taken of the differences of age in early stage learners of English and how this will affect their use of English? Is this reflected in how the frameworks are constructed?
- Does the framework accommodate evidence of pupils' unsuccessful attempts at language use as part of the overall assessment of progress?
- Does the framework provide an orientation for learners and teachers that stimulates the development of learning autonomy?

Language focus:

- Does the framework distinguish between informal use of English (social interaction) and formal use (as in some formal teaching in school contexts)?
- What view of vocabulary and grammar is adopted (functional or formal approaches)?
- Are grammar points prioritised, and if so on what basis?
- Are issues related to social and cultural conventions of language use, e.g. politeness and formality, taken into account?
- Does the framework refer to different registers and styles of English language use?
- Does the framework draw attention to the differences between spoken and written English used for different social and academic purposes?
- Does the framework refer to different genres, e.g. related to subject-specific discourse?

Language across the curriculum

- Are there guidelines for assessing English in the context of different subject disciplines, e.g. on distinguishing between language-based and subject-related cognitive difficulties?
- How are the linguistic aspects of curriculum subjects recognised, e.g. in terms of vocabulary, grammar, functional use, or specific classroom interactions?
- Is there guidance on the development of the different language skills in the context of subject-related language use?

Supporting materials and guidelines

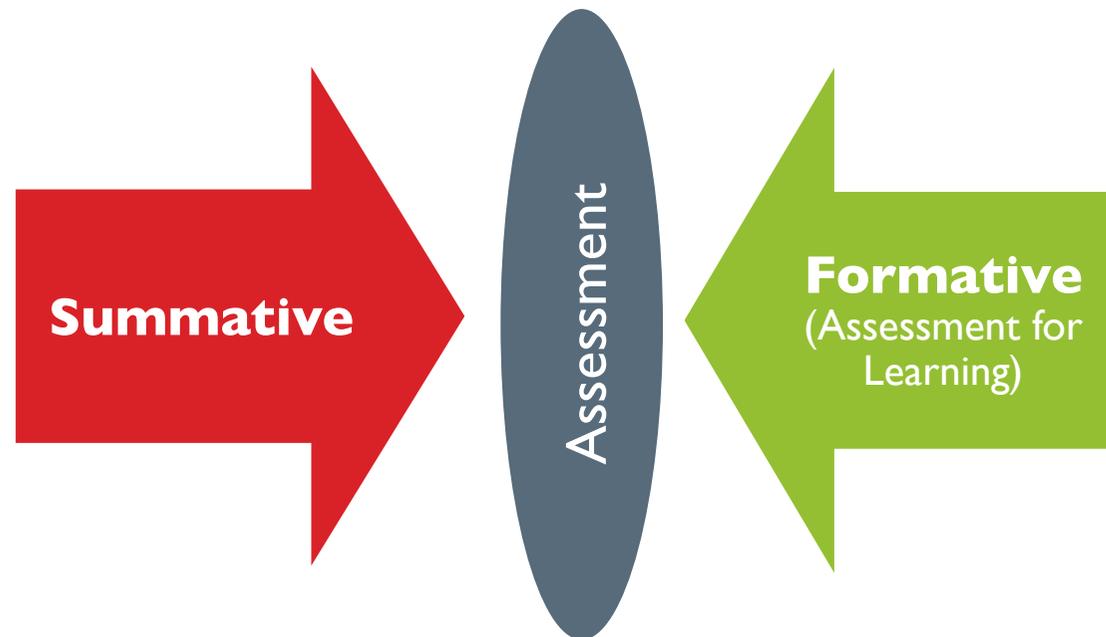
- Is there guidance for teachers on the use of the rating scales, for instance, on avoiding an excessively rigid and linear view of progression?
- Are there guidelines for assessing English in the context of different subject disciplines in the curriculum?
- Are there useful exemplifications of evidence of EAL linguistic or social classroom behaviour, which can be indicative of different stages of learning and language competence?

Assessment affects the quality of teaching and learning – what is its purpose and why do we need it?

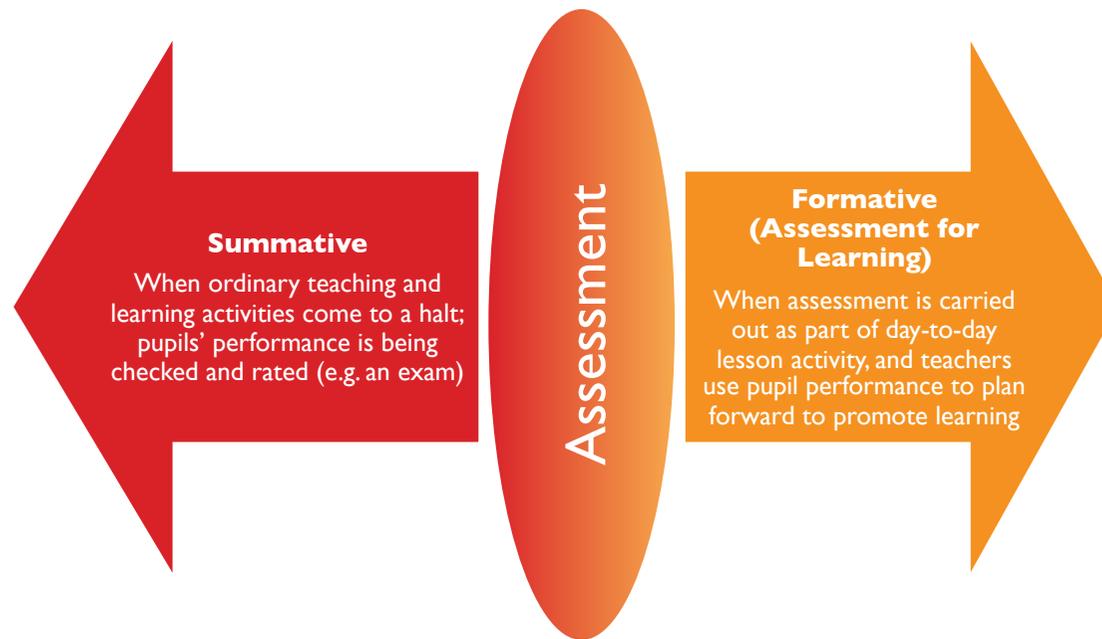
Purposes and uses of assessment

Assessment is an important part of educational provision. High quality assessment can help teachers and pupils to find out what has been achieved and what more needs to be done. While it is important for teachers to take account of the evidence of the development of EAL pupils' English language proficiency through observation of their use of the language in lessons as well as in informal interactions outside lessons, in this assessment framework we are mainly referring to assessment through teaching-learning activities.

Traditionally, assessment has been divided into two kinds: summative and formative. However, we have increasingly come to understand that assessment activities in themselves are essentially purpose-neutral.



It is the way/s in which we make use of the assessment process and outcome that would render them purpose-bound.



Validity of assessment

An important characteristic of any high quality assessment is that it is valid. What makes any assessment valid? There are three key considerations:

- whether an assessment taps into the knowledge and skills that it claims to be focusing on
- what interpretation and use is made of the assessment outcomes
- what consequences the assessment may have on the key stakeholders (e.g. pupils, teacher, parents). (Assessment Reform Group, 2002; Bachman and Palmer, 2010; Black and Wiliam, 2009; Jones and Saville, 2016; Leung 2013; Messick, 1989)

Teachers know that any assessment can only provide a snapshot of what pupils have learned and can do. So it is very important that assessment tasks (e.g. questions in an examination or a classroom quiz) are closely matched to the knowledge and skills and/or pupil abilities that we claim to be assessing.

Outcomes of assessment have to be interpreted in relation to context and pupils' background experiences and present learning trajectories. This is a particularly important issue in EAL assessment. For instance, an accurate English language Proficiency assessment of a 10-year old pupil new to English, who has only been in school for a few weeks will, in all probability, show that s/he does not have the same knowledge and skills in English as her/his English-speaking background peers. Sensitive professional interpretation of such an assessment outcome will need to take account of the pupil's previous educational experience, first language and literacy development, present disposition in relation to learning and engagement with school life generally, and the processes of additional language development in school contexts. A low assessment outcome does not mean low capacity and motivation to learn and develop in the future. So the assessment outcome provides an informed basis for making appropriate teaching provision in response to pupil needs.

It follows then that the educational decisions we make on the basis of summative and formative assessment outcomes have an ethical dimension. For instance, we may choose to use assessment outcomes to select pupils for particular pathways in the curriculum, and/or to create different teaching programmes for pupils at different levels of attainment. Teachers and schools are encouraged to take account of all three considerations set out above when conducting EAL and other assessments.

References

- Andrich, D. (1978). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, 2, 449-460.
- Arnot, M., Schneider, C., Evans, M., Liu, Y., & Welply, O., Davies-Tutt, D. (2014). *School approaches to the education of EAL students: Language development, social integration and achievement*. Cambridge: The Bell Foundation.
- Assessment Reform Group. (2002). Assessment for learning: 10 principles. Retrieved 2014-04-01, from http://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf
- Bachman, L., & Palmer, A. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5-31.
- Bramley, T. (2005). A Rank-Ordering Method for Equating Tests by Expert Judgment. *Journal of Applied Measurement* 6(2), 202-223
- DfE, (2016). School census 2016 to 2017 Guide, version 1.1, July.
- Evans, M., Schneider, C., Arnot, M., Fisher, L., Forbes, K., Hu, M., Liu, Y. (2016). *Language development and school achievement: Opportunities and challenges in the education of EAL students*. Cambridge: The Bell Foundation.
- Jones, N. (2009). A comparative approach to constructing a multilingual proficiency framework: constraining the role of standard setting. *Research Notes Issue 37* August 2009. Available from: <http://www.cambridgeenglish.org/images/23156-research-notes-37.pdf>
- Jones, N., & Saville, N. (2016). *Learning oriented assessment*. Cambridge: Cambridge University Press.
- Laming, D. (2004). *Human judgment: The eye of the beholder*. London: Thomson.
- Leung, C. (2013). Classroom-based Assessment: Issues for Language Teacher Education. In A. Kunnan (Ed.), *Companion to Language Assessment* (Vol. III, pp. 1510-1519). Hoboken, NJ: Wiley -Blackwell.
- Linacre, J. M. (2006). Rasch Analysis of Rank-Ordered Data. *Journal of Applied Measurement*, 7(11), 129-139
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd edition., pp. 13-103). New York: ACE-NCME, MacMillan.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 3, 273-286