NEAS Test of English

Construct Validity Statement

Table of Contents

Executive Summary

The NEAS Test of English has been developed using a **construct-based approach to language assessment**, grounded in internationally recognised frameworks including the **CEFR** and established research in language testing and cognitive processing.

The test assesses English proficiency as a **unified construct of communicative ability**, encompassing both:

- **comprehension** (reading and listening), and
- **production** (writing and speaking).

Across all four skills, proficiency is defined not as knowledge of language forms in isolation, but as the ability to:

construct, interpret, and communicate meaning across increasing levels of discourse complexity, abstraction, and cognitive demand.

Key features supporting construct validity

1. Clear construct definition

- The test is based on a multi-component model of language ability, including:
  - comprehension,
  - inference,
  - discourse processing,
  - and evaluative meaning.
- This aligns directly with CEFR descriptors and contemporary language assessment theory.

2. Direct alignment between construct, tasks, and scoring

- Test tasks are explicitly designed to elicit the skills defined in the construct
- Marking descriptors (for speaking and writing) define exactly what successful performance looks like
- Scores are awarded based on **observable evidence of these abilities**, not subjective impressions

3. Systematic progression across CEFR levels (A2–C1)

- Difficulty increases through:
  - greater cognitive demand,
  - more complex discourse,
  - and higher levels of abstraction
- This progression is consistent across all four skills

4. Construct-driven difficulty (not artificial difficulty)

- Difficulty is controlled through:
  - what candidates must *do* with language (e.g. infer, explain, evaluate)
- Not through:
  - trick questions,
  - obscure vocabulary,

- or arbitrary scoring

5. Strong control of fairness and bias

- Topics are culturally neutral and globally accessible
- Performance is assessed based on **how language is used**, not background knowledge
- Listening conditions (e.g. speed, length, accent) are carefully controlled

6. Explicit operationalisation of the construct in productive skills

- Writing and speaking are assessed using **descriptor-based criteria**
- These criteria are linked to:
  - observable performance features
  - CEFR descriptors
  - and task demands
- Raters award scores based on **consistent evidence of ability**, ensuring reliability and transparency

Conclusion

The NEAS Test of English demonstrates **strong construct validity at the design level**. The test:

- clearly defines the construct of language proficiency,
- consistently represents that construct across all skills,
- aligns tasks and scoring with the intended abilities,
- and ensures that performance reflects genuine language use.

While statistical validation will follow through pilot testing, the current design provides a **coherent, theoretically grounded, and defensible assessment framework** aligned with international best practice.

Full Construct Validity Statement

1. Purpose of this Document

This document presents the **construct validity argument** for the NEAS Test of English. It demonstrates how the test:

- defines communicative language proficiency,
- represents that construct across all skills,
- and operationalises it through task design and scoring.

In line with contemporary validity theory, validity is understood not as a property of the test itself, but as:

an evidence-based argument supporting the interpretation of test scores as indicators of language ability.

Accordingly, this document establishes that:

> *scores from the NEAS Test of English can be interpreted as meaningful and appropriate representations of communicative language proficiency, aligned with the CEFR.*

2. Definition of the Construct

The NEAS Test of English is based on a **communicative, CEFR-aligned model of language ability**, in which proficiency is defined as:

the ability to interpret, construct, and communicate meaning across a range of contexts, with increasing control over discourse, abstraction, and linguistic resources.

Language ability is therefore conceptualised as:

- performance in communicative tasks,
- not the possession of discrete linguistic knowledge.

The construct integrates:

- comprehension (reading, listening),
- production (writing, speaking),
- discourse processing,
- inferencing and meaning development,
- and the use of linguistic resources as tools for communication.

This definition aligns with:

- CEFR (action-oriented model),
- communicative competence frameworks (Bachman; Canale & Swain),
- and socio-cognitive validation models (Weir).

3. Construct Representation Across Skills

3.1 Receptive Skills (Reading and Listening)

Receptive skills assess the ability to **construct meaning from input**, including:

- global understanding,
- detail recognition,
- inferencing,
- discourse interpretation,
- and evaluation at higher levels.

Tasks are designed to elicit these processes explicitly, ensuring that:

> *the test samples the full construct of comprehension rather than isolated subskills.*

3.2 Productive Skills (Writing and Speaking)

Productive skills assess the ability to **generate and communicate meaning**, requiring candidates to:

- organise and develop ideas,
- manage discourse,
- and use language resources effectively.

Crucially, the construct is realised through:

- **task demands**, and
- **marking descriptors**, which define what constitutes successful performance.

4. Descriptors as Measurement Instruments

In the productive skills, marking descriptors are not treated as informal guidance, but as:

central measurement instruments that operationalise the construct.

Descriptors:

- define the observable features of proficiency,
- establish how performance is interpreted,
- and provide the basis for scoring decisions.

They are:

- behaviourally anchored,
- aligned with CEFR descriptors,
- and structured to reflect developmental progression.

This ensures that:

> *the construct is not only defined theoretically but realised in measurable and observable terms.*

5. Observable Performance and Operationalisation

All scoring criteria are grounded in **observable features of performance**, including:

- coherence → logical sequencing and cohesion
- development → explanation, support, evaluation
- lexical resource → range and appropriacy
- grammar → control and impact on meaning
- fluency → continuity and ease of delivery

This emphasis on observability ensures:

- transparency in scoring,
- consistency across raters and systems,
- and compatibility with both human and AI-mediated assessment.

6. Scoring Philosophy and Interpretation of Error

A central principle of the assessment is that:

proficiency is defined by control appropriate to level, not by absence of error.

Errors are evaluated:

- in terms of their **impact on meaning**,
- not simply their frequency or form.

Accordingly:

- minor errors that do not affect comprehension are tolerated at higher levels
- more serious errors are interpreted based on their effect on communicative effectiveness

This reflects:

- CEFR principles
- and research on accuracy–complexity trade-offs

and ensures that:

> *scoring reflects communicative ability rather than formal correctness alone.*

7. Alignment Between Construct, Tasks, and Scoring

The NEAS test demonstrates a high degree of alignment between:

Component	Function
Construct	Defines ability
Tasks	Elicit performance
Descriptors	Define evidence
Scores	Reflect observed performance

This alignment ensures that:

> *test results are valid representations of the intended construct.*

8. Progression and CEFR Alignment

Proficiency is structured as a developmental continuum:

Level	Characteristics
A2	Basic, concrete communication
B1	Connected meaning, emerging discourse
B2	Structured explanation and reasoning
C1	Abstract, evaluative, flexible communication

Progression is defined by:

- increasing cognitive demand
- discourse complexity
- abstraction
- and control of language

This progression is consistently realised across:

- tasks
- texts
- and scoring criteria

9. Construct-Driven Difficulty

Difficulty is determined by:

- processing demands,
- discourse requirements,
- and communicative complexity

—not by:

- superficial features such as text length or vocabulary alone.

This ensures that:

> *higher scores reflect higher levels of language ability, not artefacts of test design.*

10. Sampling and Construct Coverage

The test samples performance across a **range of CEFR levels**, particularly in receptive skills.

This design:

- reduces ceiling and floor effects
- improves classification accuracy
- and ensures broad construct coverage

As a result:

> *proficiency is inferred from patterns of performance across the construct, rather than single-task outcomes.*

11. Rater Decision Framework

Scoring decisions are based on:

consistent evidence of performance across criteria

For productive skills:

- levels are awarded when performance is stable across dimensions
- isolated higher-level features are not sufficient

This ensures:

- reliability of scoring
- and alignment with CEFR definitions of proficiency

12. Control of Construct-Irrelevant Variance

The test minimises external influences on performance, including:

- cultural or topic bias
- background knowledge
- accent familiarity
- test-taking strategies

Topics are:

- generalisable
- culturally neutral
- and accessible

This ensures:

> *scores reflect language ability rather than extraneous factors.*

13. AI-Mediated Design and Validity

The AI-mediated delivery model introduces constraints that have been incorporated into the validity framework.

In particular, it requires that:

- performance features are observable and measurable

- scoring criteria are clearly defined and consistent

As a result:

the construct has been operationalised in a way that supports both human and automated scoring without compromising communicative validity.

Rather than weakening the construct, this has:

- strengthened clarity of definition
- and enhanced consistency of measurement

14. Theoretical Foundation

The test is grounded in:

- CEFR (Council of Europe)
- Messick's unified theory of validity
- Weir's socio-cognitive model
- Bachman & Palmer's communicative language ability
- established research in reading, listening, speaking, and writing assessment

15. Conclusion

The NEAS Test of English demonstrates strong construct validity through:

- a clearly defined and theoretically grounded construct,
- consistent representation across all skills,
- explicit alignment between tasks and scoring,
- operationalisation through observable performance features,
- and a coherent, CEFR-aligned progression of proficiency.

The test therefore provides:

> *a defensible, transparent, and internationally aligned measure of communicative language ability.*