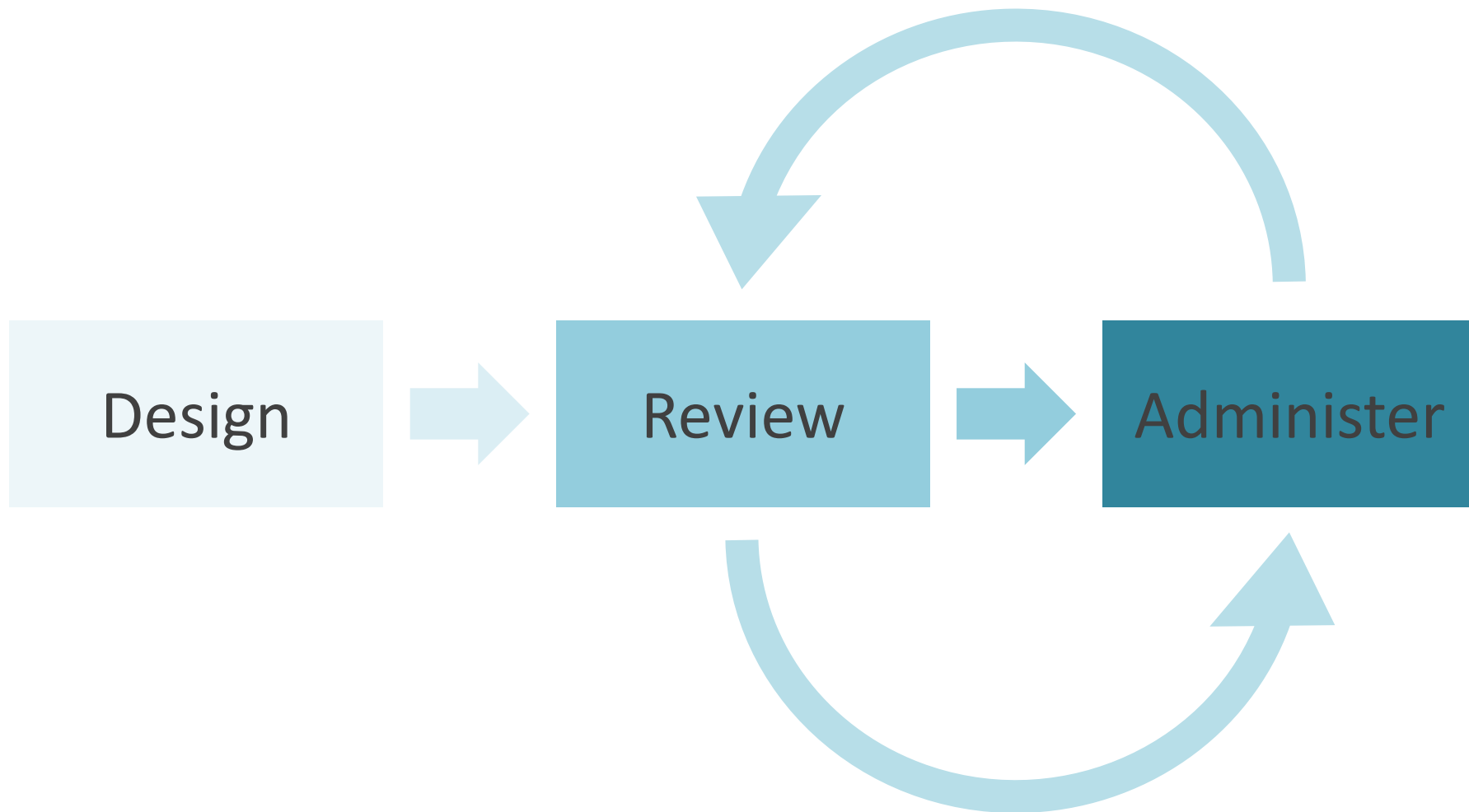
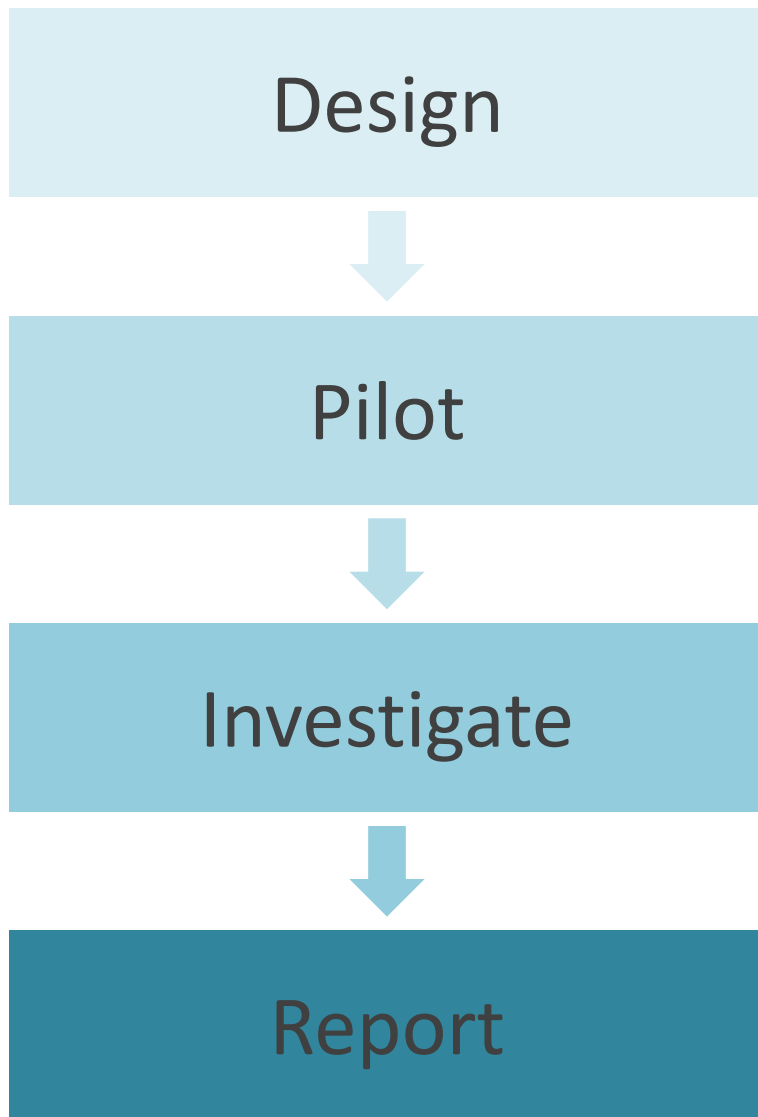




Designing, evaluating and improving language tests

Henry Emery





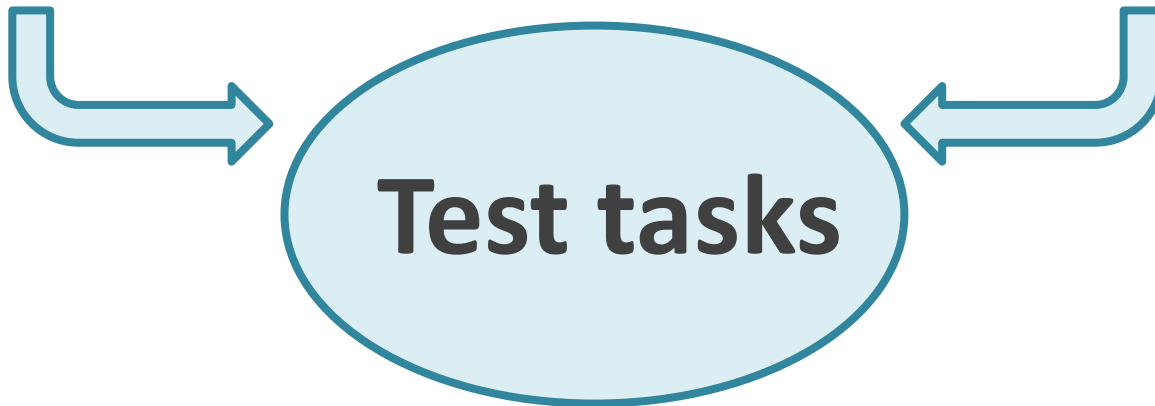
Design

- Who are we testing?
- Do our tasks together represent:
 - On-the-job language use?
 - The six criteria?
 - The six levels?
- How is the test assessed?

- ground, approach, tower and en-route ATC
- flight crew, cabin crew and passengers
- maintenance personnel
- operations and dispatch
- tug drivers, de-icing crews etc
- taxi drivers, hotel staff...
- simulator instructors and examiners
- *language testers?*

On-the-job language use tasks

Level	Pronunciation	Structure	Vocabulary	Fluency	Comprehension	Interactions
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Ability to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations, is sensitive to verbal and non-verbal cues, and responds to them appropriately.
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, seldom interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work related topics. Paragraphs consistently and successfully. Vocabulary is sometimes idiomatic.	Ability to speak at length with relative ease on familiar topics, but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
Optional 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous speech. When the speaker is confronted with a linguistic or situational complication, can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.
Optional 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	Basic grammatical structures and sentence patterns associated with predictable situations are not always well controlled. Errors frequently interfere with meaning.	Vocabulary range and accuracy are often sufficient to communicate on common, concrete, or work related topics but the range is limited and the word choice often inappropriate. Is often unable to paraphrase successfully when lacking vocabulary.	Produces stretches of language, but phrasing and pacing are often inappropriate. Hesitations or slowness in language processing may prevent effective communication. Fillers are sometimes distracting.	Comprehension is often accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. May fail to understand a linguistic or situational turn of events.	Responses are sometimes immediate, appropriate, and informative. Can initiate and maintain exchanges with reasonable ease on familiar topics and in predictable situations. Generally inadequate when dealing with an unexpected turn of events.
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	Shows only limited control of a few simple memorized grammatical structures and sentence patterns.	Limited vocabulary range consisting only of isolated words and memorized phrases.	Can produce very short, isolated, memorized utterances with frequent pausing and a distracting use of fillers to search for expressions and to articulate less familiar words.	Comprehension is limited to isolated, memorized phrases when they are carefully and slowly articulated.	Response time is slow, and often inappropriate. Interaction is limited to simple routine exchanges.
Below Elementary 1	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.



Subject matter
knowledge



Language
knowledge

Subject matter knowledge

Scenario:

EFOT, AC climbing to 5,000 ft. Intend to return to airfield.

Task:

Dump fuel

English language knowledge

C: 321 heavy. And will you be dumping fuel?

P: ***321 heavy we will have... yes... maybe three minutes' dumping***

C: OK, do you want to do that over the ocean right now that would be probably the best bet

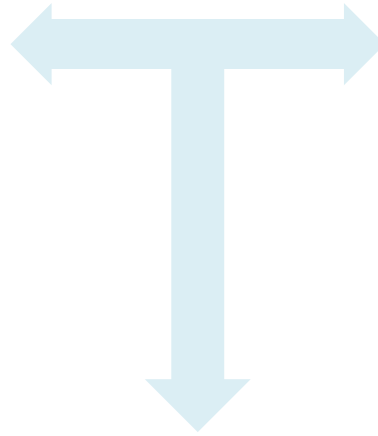
P: ***That is affirmative. I'll call you back when starting if it is possible on that heading***

C: 321 heavy. If possible fly heading one eight zero and I'll just keep you over the water while you do the dump

P: ***OK, maintain ... heading one eight zero call you back ending... uh... dumping***



Subject
matter
knowledge



Language
knowledge

Test tasks

Level	Pronunciation	Structure	Vocabulary	Fluency	Comprehension	Interactions
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues, and responds to them appropriately.
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics, but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.
Pre-Operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	Basic grammatical structures and sentence patterns associated with predictable situations are not always well controlled. Errors frequently interfere with meaning.	Vocabulary range and accuracy are often sufficient to communicate on common, concrete, or work related topics but range is limited and the word choice often inappropriate. Is often unable to paraphrase successfully when lacking vocabulary.	Produces stretches of language, but phrasing and pausing are often inappropriate. Hesitations or slowness in language processing may prevent effective communication. Fillers are sometimes distracting.	Comprehension is often accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. May fail to understand a linguistic or situational turn of events.	Responses are sometimes immediate, appropriate and informative. Can initiate and maintain exchanges with reasonable ease on familiar topics and in predictable situations. Generally inadequate when dealing with an unexpected turn of events.
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	Shows only limited control of a few simple memorized grammatical structures and sentence patterns.	Limited vocabulary range consisting only of isolated words and memorized phrases.	Can produce very short, isolated, memorized utterances with frequent pausing and a distracting use of fillers to search for expressions and to articulate less familiar words.	Comprehension is limited to isolated, memorized phrases when they are carefully and slowly articulated.	Response time is slow, and often inappropriate. Interaction is limited to simple routine exchanges.
Pre-Elementary 1	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.

...a wide variety of familiar and unfamiliar topics...

...comprehension of linguistic and cultural subtleties...

...transition from rehearsed or formulaic speech to spontaneous interaction...

...is sensitive to non-verbal cues...

Level	Pronunciation	Structure	Vocabulary	Fluency	Comprehension	Interactions
	Assumes a dialect and/or accent intelligible to the aeronautical community.	Relevant grammatical structures and sentence patterns are determined by language functions appropriate to the task.				
Expert 6	Pronunciation, stress, rhythm, and intonation, though possibly influenced by the first language or regional variation, almost never interfere with ease of understanding.	Both basic and complex grammatical structures and sentence patterns are consistently well controlled.	Vocabulary range and accuracy are sufficient to communicate effectively on a wide variety of familiar and unfamiliar topics. Vocabulary is idiomatic, nuanced, and sensitive to register.	Able to speak at length with a natural, effortless flow. Varies speech flow for stylistic effect, e.g. to emphasize a point. Uses appropriate discourse markers and connectors spontaneously.	Comprehension is consistently accurate in nearly all contexts and includes comprehension of linguistic and cultural subtleties.	Interacts with ease in nearly all situations. Is sensitive to verbal and non-verbal cues, and responds to them appropriately.
Extended 5	Pronunciation, stress, rhythm, and intonation, though influenced by the first language or regional variation, rarely interfere with ease of understanding.	Basic grammatical structures and sentence patterns are consistently well controlled. Complex structures are attempted but with errors which sometimes interfere with meaning.	Vocabulary range and accuracy are sufficient to communicate effectively on common, concrete, and work related topics. Paraphrases consistently and successfully. Vocabulary is sometimes idiomatic.	Able to speak at length with relative ease on familiar topics, but may not vary speech flow as a stylistic device. Can make use of appropriate discourse markers or connectors.	Comprehension is accurate on common, concrete, and work related topics and mostly accurate when the speaker is confronted with a linguistic or situational complication or an unexpected turn of events. Is able to comprehend a range of speech varieties (dialect and/or accent) or registers.	Responses are immediate, appropriate, and informative. Manages the speaker/listener relationship effectively.
Operational 4	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation but only sometimes interfere with ease of understanding.	Basic grammatical structures and sentence patterns are used creatively and are usually well controlled. Errors may occur, particularly in unusual or unexpected circumstances, but rarely interfere with meaning.	Vocabulary range and accuracy are usually sufficient to communicate effectively on common, concrete, and work related topics. Can often paraphrase successfully when lacking vocabulary in unusual or unexpected circumstances.	Produces stretches of language at an appropriate tempo. There may be occasional loss of fluency on transition from rehearsed or formulaic speech to spontaneous interaction, but this does not prevent effective communication. Can make limited use of discourse markers or connectors. Fillers are not distracting.	Comprehension is mostly accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. When the speaker is confronted with a linguistic or situational complication or an unexpected turn of events, comprehension may be slower or require clarification strategies.	Responses are usually immediate, appropriate, and informative. Initiates and maintains exchanges even when dealing with an unexpected turn of events. Deals adequately with apparent misunderstandings by checking, confirming, or clarifying.
Pre-Operational 3	Pronunciation, stress, rhythm, and intonation are influenced by the first language or regional variation and frequently interfere with ease of understanding.	Basic grammatical structures and sentence patterns associated with predictable situations are not always well controlled. Errors frequently interfere with meaning.	Vocabulary range and accuracy are often sufficient to communicate on common, concrete, or work related topics but range is limited and the word choice often inappropriate. Is often unable to paraphrase successfully when lacking vocabulary.	Produces stretches of language, but phrasing and pausing are often inappropriate. Hesitations or slowness in language processing may prevent effective communication. Fillers are sometimes distracting.	Comprehension is often accurate on common, concrete, and work related topics when the accent or variety used is sufficiently intelligible for an international community of users. May fail to understand a linguistic or situational turn of events.	Responses are sometimes immediate, appropriate, and informative. Can initiate and maintain exchanges with reasonable ease on familiar topics and in predictable situations. Generally inadequate when dealing with an unexpected turn of events.
Elementary 2	Pronunciation, stress, rhythm, and intonation are heavily influenced by the first language or regional variation and usually interfere with ease of understanding.	Shows only limited control of a few simple memorized grammatical structures and sentence patterns.	Limited vocabulary range consisting only of isolated words and memorized phrases.	Can produce very short, isolated, memorized utterances with frequent pausing and a distracting use of fillers to search for expressions and to articulate less familiar words.	Comprehension is limited to isolated, memorized phrases when they are carefully and slowly articulated.	Response time is slow, and often inappropriate. Interaction is limited to simple routine exchanges.
Pre-Elementary 1	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.	Performs at a level below the Elementary level.

Pilot

- Do the rubrics / instructions work?
- Do the tasks work?
- Does the timing work?
- Do the procedures work?
- Does the test *flow*?

Investigate

- How difficult are the tasks?
- Are the items (questions) / tasks working well?
- Are versions of the test equally difficult?
- Do the tasks / raters produce reliable scores?
- What do independent experts think?
- What do the candidates think?

Investigate

**WHAT IMPROVEMENTS
CAN WE MAKE?**

Report

- Information for candidates
- Information for other stakeholders
- **Evidence** for test quality
- **Plans** for improvement

Administer
Test v1



Investigate

- Are the systems and procedures working well?
- Is the test producing consistent results?
- Is the test discriminating at the various levels?
- Are the versions equally difficult?
- Is the test biased towards one group or another?

Administer
Test v1

Investigate

Review

Investigate

Report



Administer
Test v2

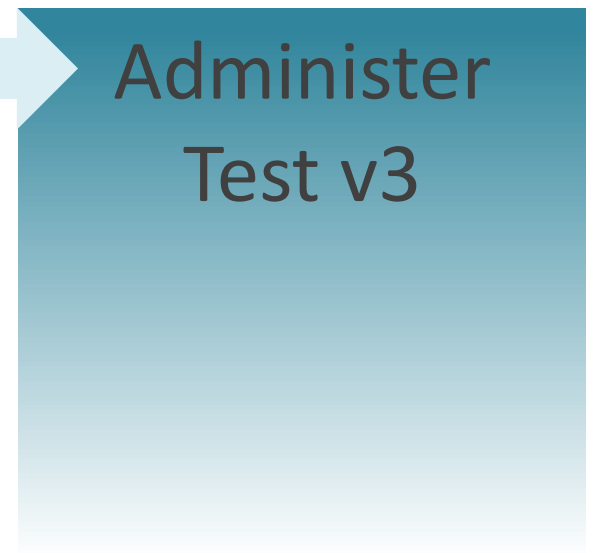
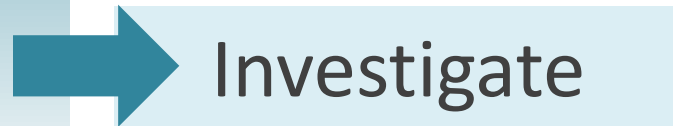
Administer
Test v2



Investigate

- What cognitive processes do the candidates apply?
- What cognitive processes do the raters apply?
- What effect does the test have on language learning?
- Does the test produce scores consistent with other validated tests?
- **Does the test predict on-the-job language performance?**

Administer
Test v2



Of all the concepts in testing ... validity is the most basic and far-reaching, for without validity, a test ... and any inferences made from it are meaningless

Hubley & Zumbo 1996

Validity is not a characteristic of a test, but a feature of the inferences made on the basis of test scores and the uses to which a test is put

Alderson, 2002



Does it fly?

IS IT SAFE?

A photograph of a control tower and an airplane flying in a cloudy sky. The control tower is on the left, and the airplane is in the upper center. The sky is blue with white clouds.

**THANK
YOU!**

henry@latitude-aes.co.uk

References:

Alderson, J. C 2002 *Conceptions of validity and validation*. Paper presented at a conference in Bucharest, June 2002.

Hubley, A. M. & Zumbo, B. D. A dialectic on validity: where we have been and where we are going. *The Journal of General Psychology* 1996. 123[3] 207-215