


Assessment Today: The Same but Different

David Booth
Pearson English



Assessment Today – The same but different; helping teachers and learners in the new normal

- What has changed in assessment in the last 5 years
- Education technology and modern language assessment practice
- Why and how do we use the technology?
 - Automated scoring and integrated skills
 - Remote proctored testing
- Linking assessment and learning in the classroom
- Questions



Educational Technology – Now and then.







Question

Think about your own day-to-day lives. To what extent do you make use of Artificial Intelligence in your daily lives?



Why do we use computers in assessment?

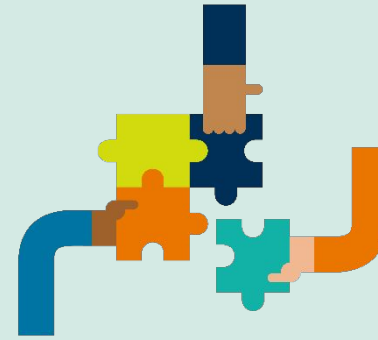




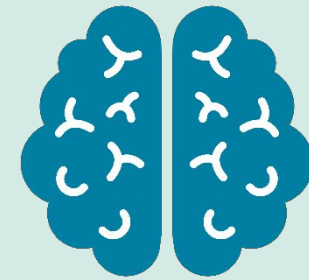
Quicker



Able to do routine
tasks without
getting bored



Make fewer
mistakes



Make complex
tasks easier

Calculation of ability and variance in adaptive testing

Computer expected score and variance for M:

For each item i of difficulty D_i , the probability of person n 's success on item i
 $= P_i = 1 / (1 + e^{(D_i - M)})$

where $e = 2.7183$

person n 's total raw score
 $= \textbf{Score} = \Sigma(P_i) \text{ for } i=1,L$

the model variance of person n 's raw score
 $= \textbf{Variance} = \Sigma(P_i (1 - P_i)) \text{ for } i=1,L$

$$\textit{Score} = \sum_{i=1}^L \left(\frac{e^{(M-D_i)}}{1 + e^{(M-D_i)}} \right)$$

$$\textit{Variance} = \sum_{i=1}^L \left(\frac{e^{(M-D_i)}}{(1 + e^{(M-D_i)})^2} \right)$$

AI and language testing

- When we say AI what do we mean?
- Computers have been used in learning and assessment for a long time
- The difference with AI is that it brings together 3 critical elements
- Computing power, data and powerful algorithms
- And 'computers' start to learn for themselves
- This is AI

AI and Pearson

- The focus for Pearson is on the marking of speaking and written responses
- Pearson also uses adaptive algorithms in placement testing
- Pearson also uses data from testing and databases of learning objectives linked to learning material to create a virtuous circle of learning
- AI is not putting a paper and pencil test onto a screen. We need to use the power of computing to transform the way we test.



Why is this important?



Because it improves

Accuracy

Authenticity

Speed

Relevance



Automated Assessment



Question

Do you think Artificial Intelligence can be used to assess the speaking and writing language performance of language learners?

Yes

Yes, but it has limitations.

Yes, but it's not the best way.

No

Challenges for human scoring

Consistent application of the scoring rubric by many raters

Consistent application of the scoring rubric by the same rater

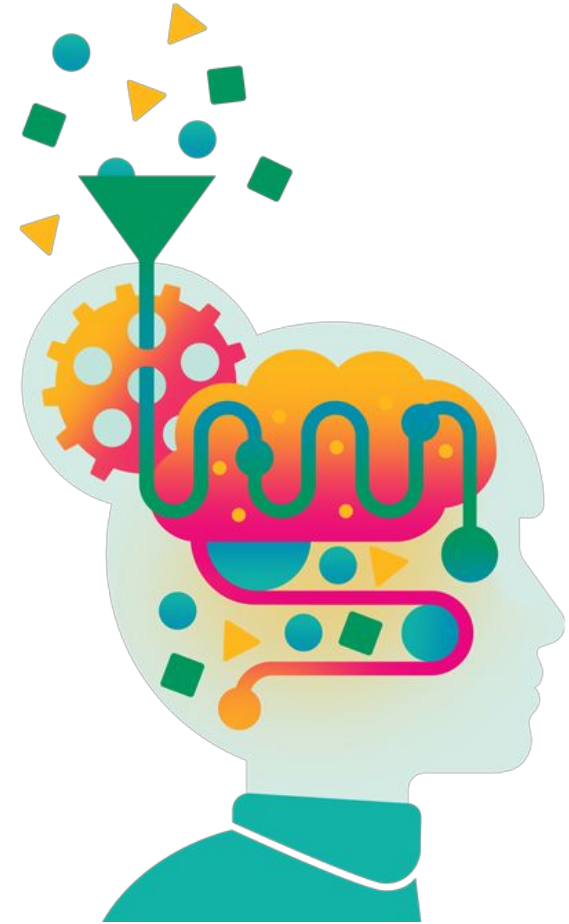
Consistency of judgments over time including time from training interventions

Use of the extreme score categories - for example the use of 1 or 2 on a 5-point scale

Cross contamination of scores for different categories or traits

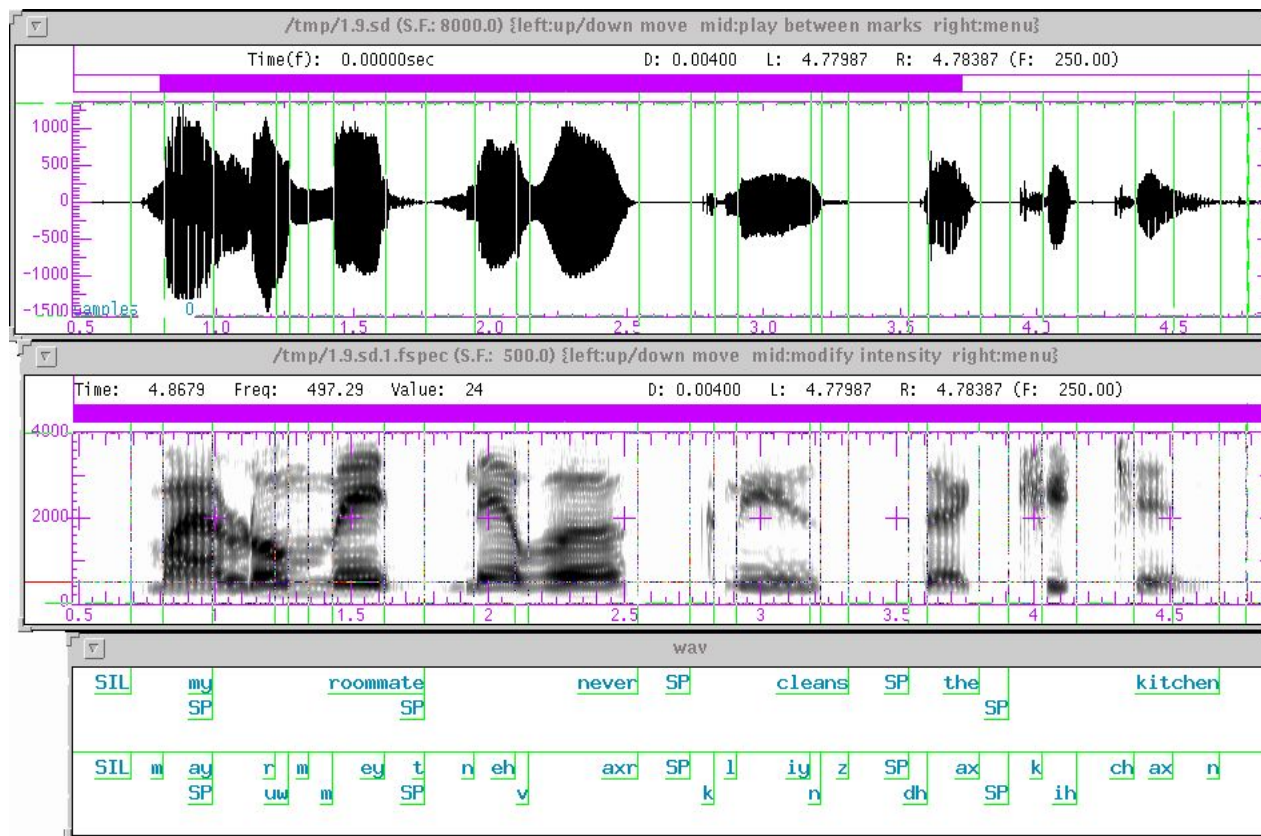
Bias due to non-linguistic features; handwriting, disagreement with ideas, tone, accent, appearance, interactional style, gender, ethnicity

Comprehension issues leading to misunderstanding and inappropriate responses



Automatic Speech Recognition

My
roommate
never cleans
the kitchen



Waveform

Spectrum

Words Segmentation

w_1 w_2 w_3 w_4 w_5 w_6

p p $pppp$ p p p p pp ppp pp p p p p p

75-90 Words/Min

5.8 Phones/Sec

The development process of Speaking Items

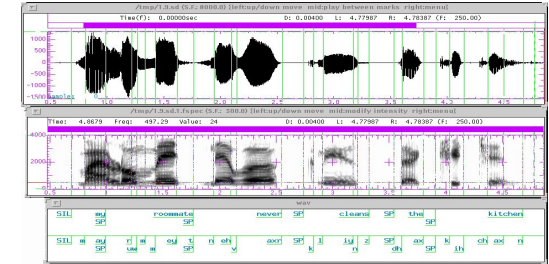
Development



Transcribing



Rating

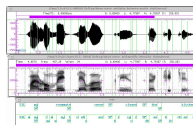


System is "trained" to predict human ratings

Validation



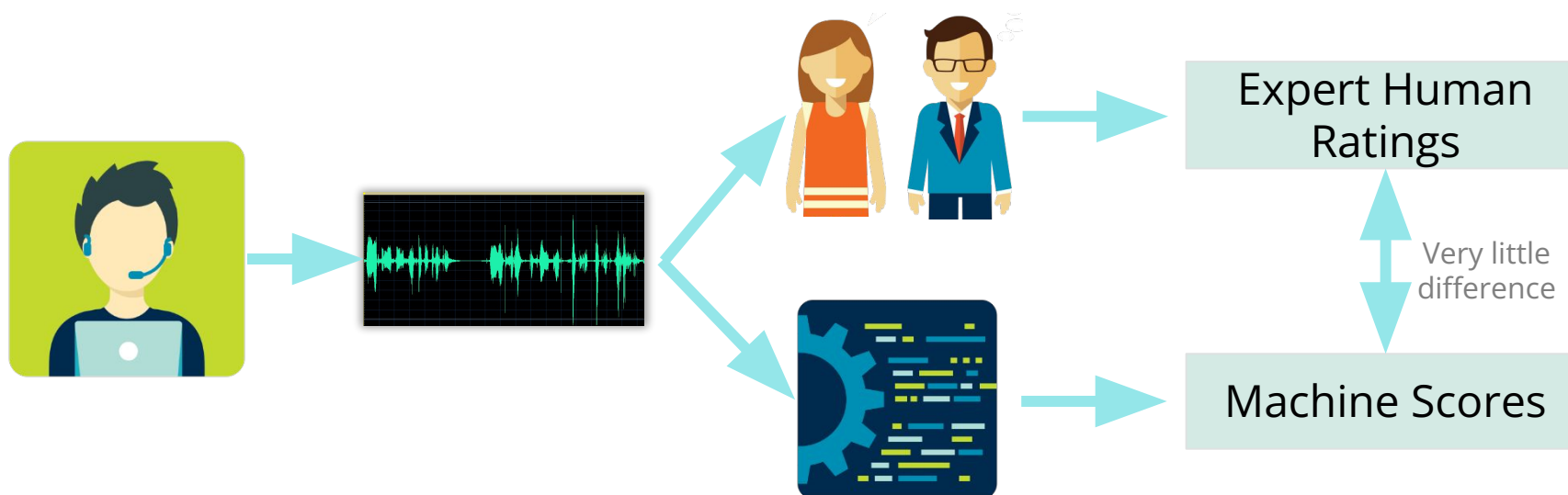
Expert human ratings



Very highly correlated

Machine scores

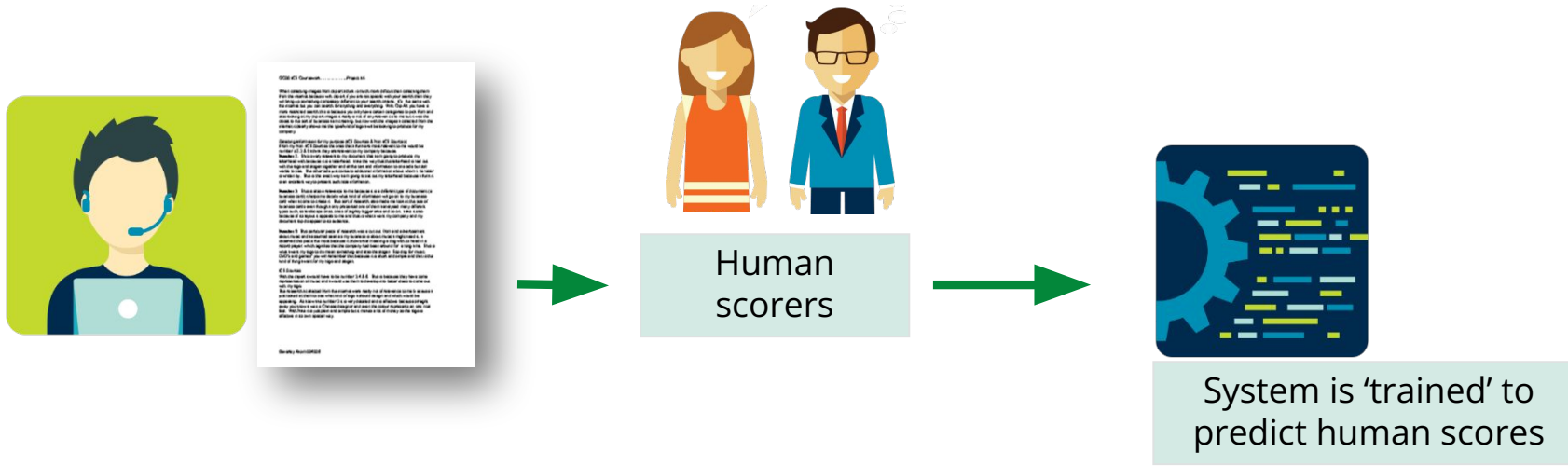
Speaking Scores: Reliability



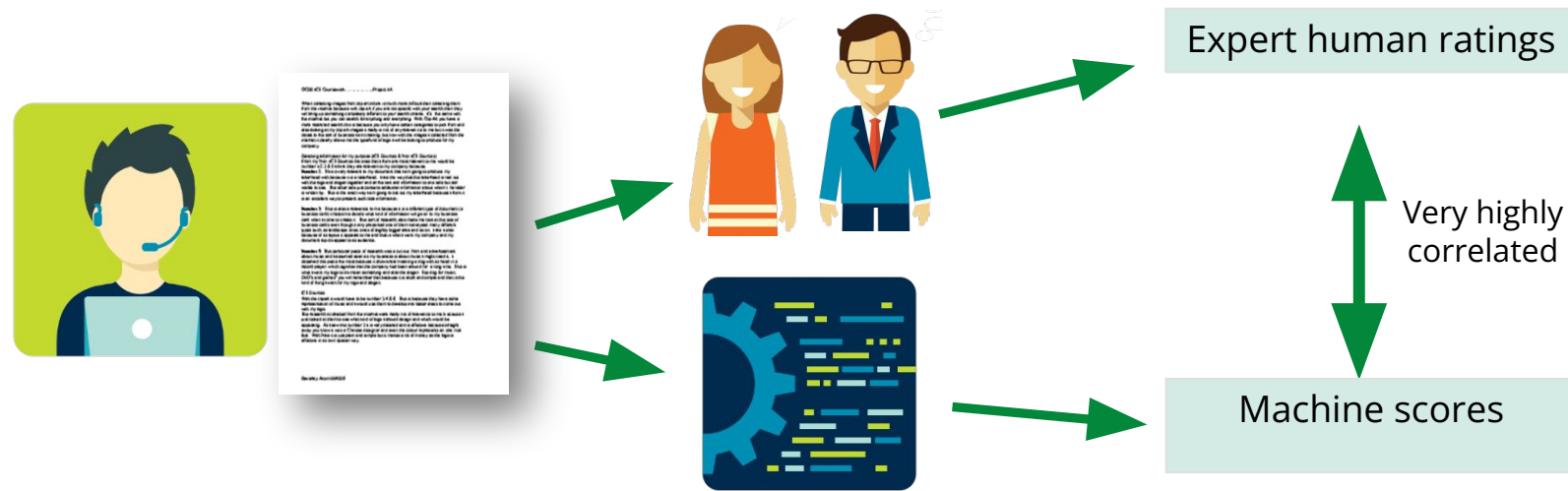
Machine-Human Correlation
(N=158)

Pronunciation	0.81
Fluency	0.82
Content	0.92
Vocabulary	0.90
Accuracy	0.95
Overall	0.96

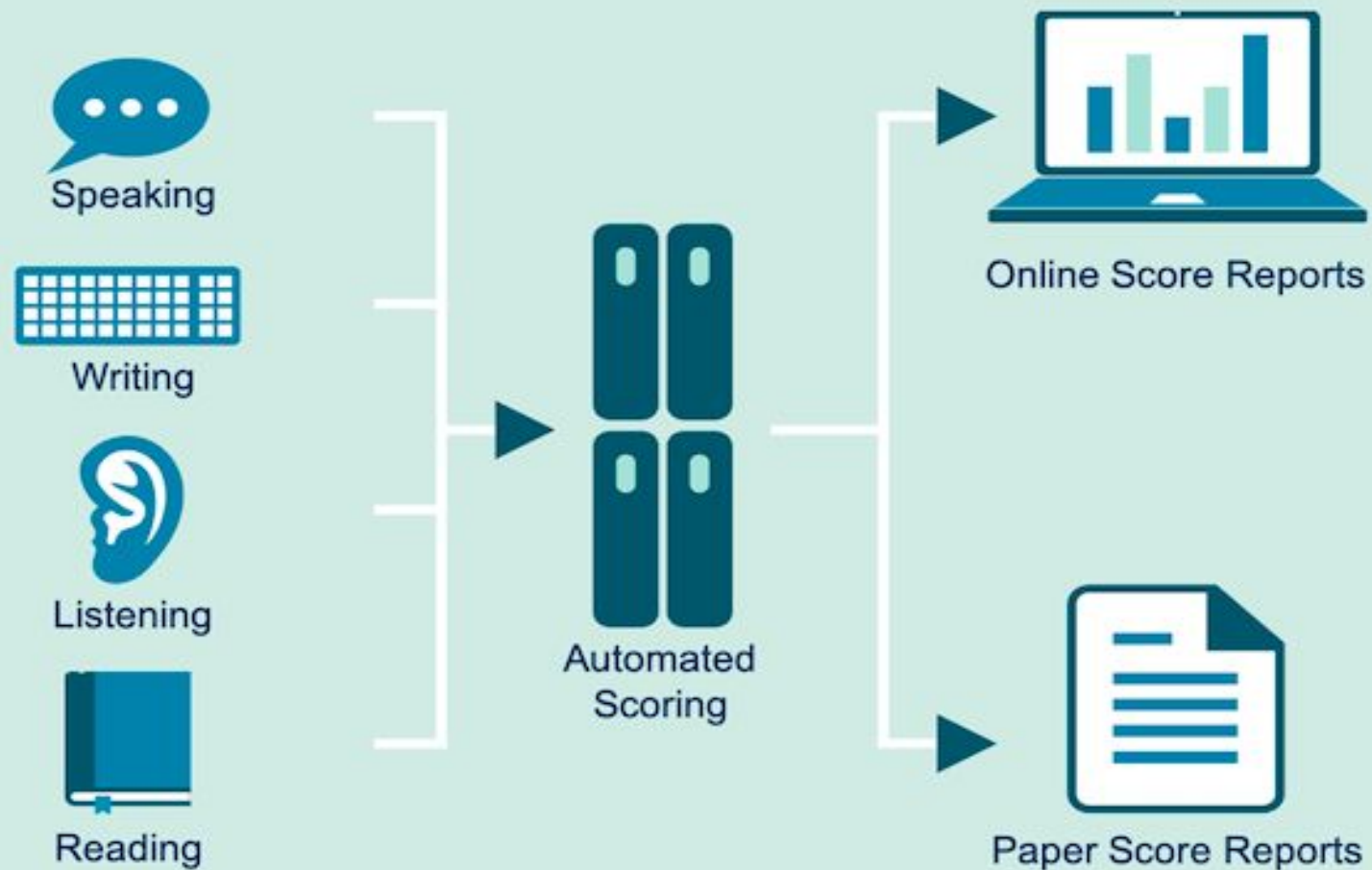
Writing Scores: Reliability



- Validation – new essays



Pearson's proprietary and patented technology drives automated assessments





Integrated Skills

Add up the following scores

Listening = 15 (6)

Reading = 22 (7.5)

Writing = 18 (6.5)

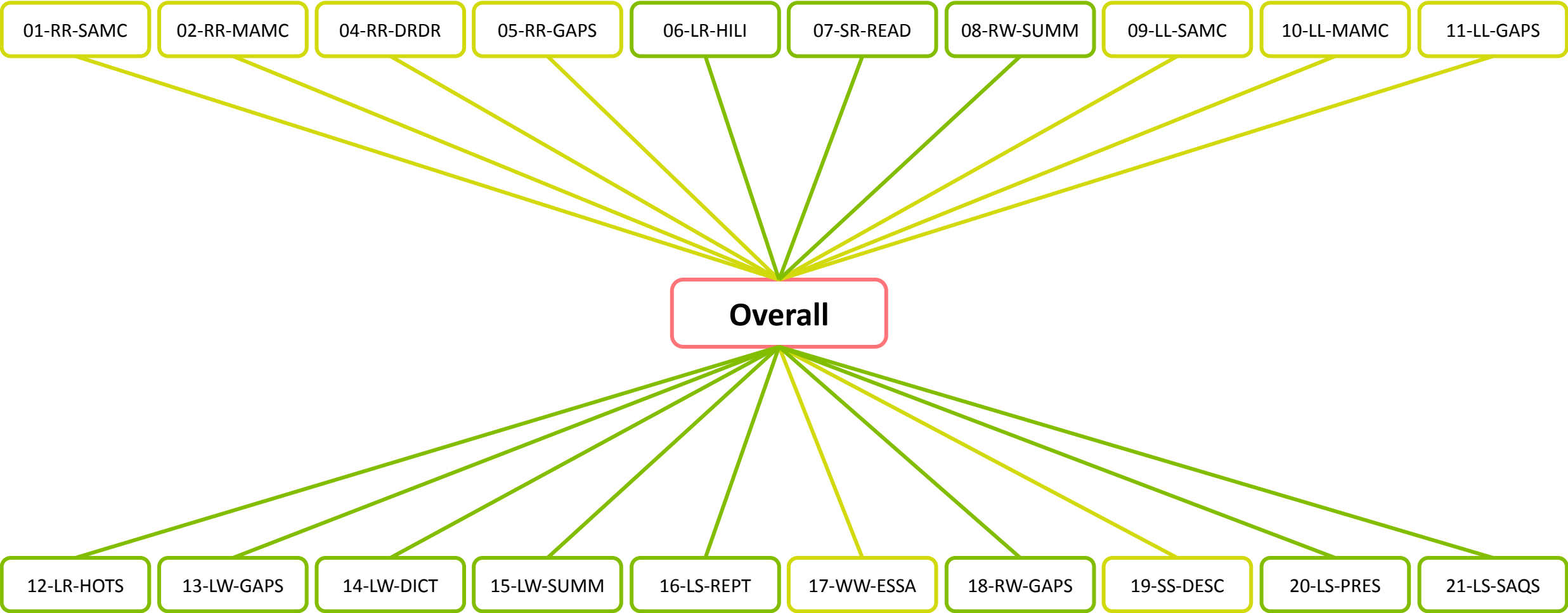
Speaking = 16 (6.5)

Total = ?? /100 (?.?)

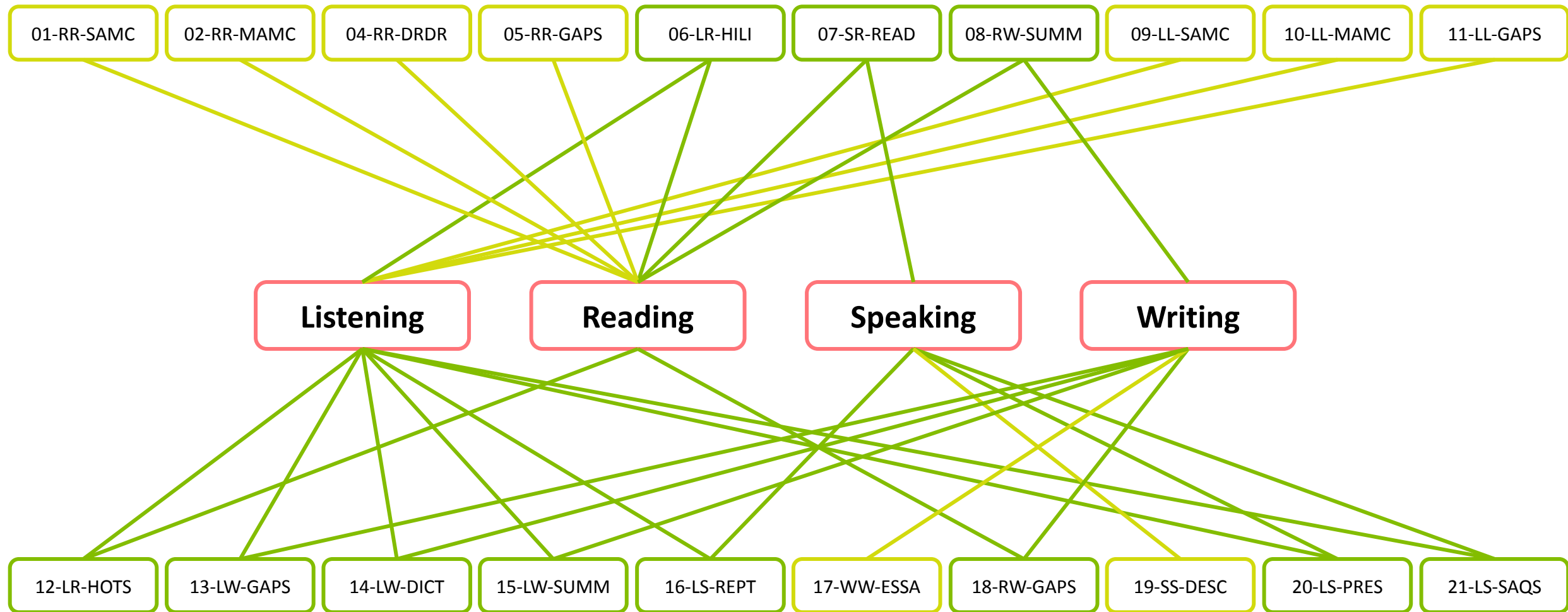
So what does that mean?



Overall Skill



Communicative Skills



AI and Educational Technology

Integrated skills testing uses all the information provided by the prospective student

- This means the assessment is **more representative** of what the student can actually do

A computer-based test gives you the opportunity to use innovative item types and more score points

- This means the test is **more accurate** than other similar tests

Automated assessment removes the impact of human bias

- This means the test **is fairer** and **more representative** of what the student can actually do

A fully automated computer test gives integrity and security in its delivery as well as accuracy of scoring

- This means you can **trust** the score you get



Remote Proctoring Solutions

PTE Academic Online



PTE Academic Online is identical to the new shortened version PTE Academic exam in content and difficulty.

Powered by **OnVue's** secure software, the test includes:

- A **check-in process** requiring test takers to submit photos of their government-issued ID. Identity checks are carried out by AI and supported by a 'greeter' who will verify the test taker's identity.
- A **live proctor** who monitors the entire test session for any suspicious behaviour - including movement outside the camera view or more than one-person present in the room.
- A **locked browser** that detects navigation away from the test window or if any new tabs are opened.
- **AI Monitoring** throughout the test which provides an additional layer of security

Security measures are applied throughout

- PTE Academic Online uses the ideal **combination of AI and human input** to deliver a best-in-class secure remote proctoring solution
- Identity checks are carried out by AI and supported by a check-in process with an OnVUE human 'greeter'
- **Human monitoring and webcam recording** occurs throughout with AI detection of unusual activity (eg multiple faces or different faces)
- The proctor can communicate with the test taker and revoke the exam in the case of **suspicious activity**
- Suspicious activity will also be reviewed after the test and **scores can be held or revoked**
- Proctors are globally based and **randomly assigned**
- There is no break in the standard test

Medium stakes testing

AI algorithms detect any suspicious activities during testing, empowering administrators to review whether a test result seems authentic.

How it works

Image monitoring: a different face or multiple faces appearing in the video frame, or camera blocked

Video monitoring: test taker moving out of camera view, more than one person in camera view, or looking away from the camera multiple times

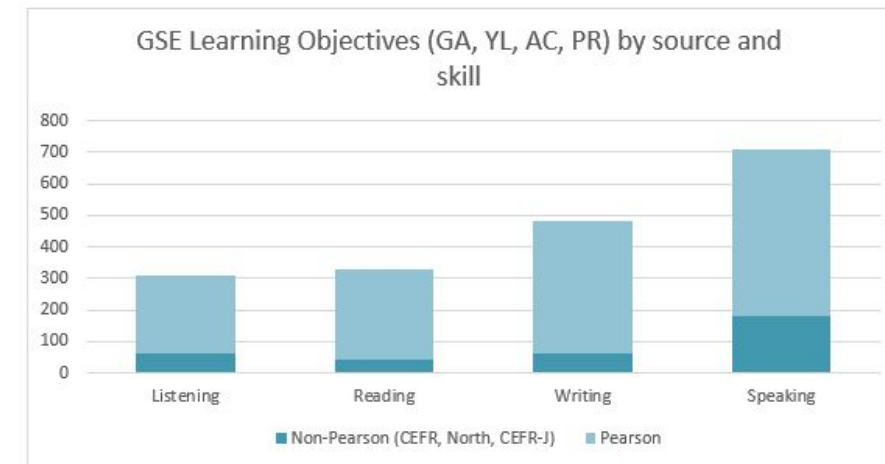
Browser monitoring: navigating away from the test window or changing tabs multiple times will be counted as violations

Complete recording: access anytime for audits and reviews. Suspicious segments highlighted for easy review



Linking Learning and Assessment

GSE Learning objectives



Global Scale of English

Sample training task

SAMPLE DESCRIPTOR

Can briefly give reasons and explanations for opinions, plans and actions.

First, decide which CEFR level you think a student would need to be at in order to have a 50% chance of successfully completing the task

Then, decide where within that level you think the learning objective sits. Use the GSE value to indicate start, middle or end

Answer:

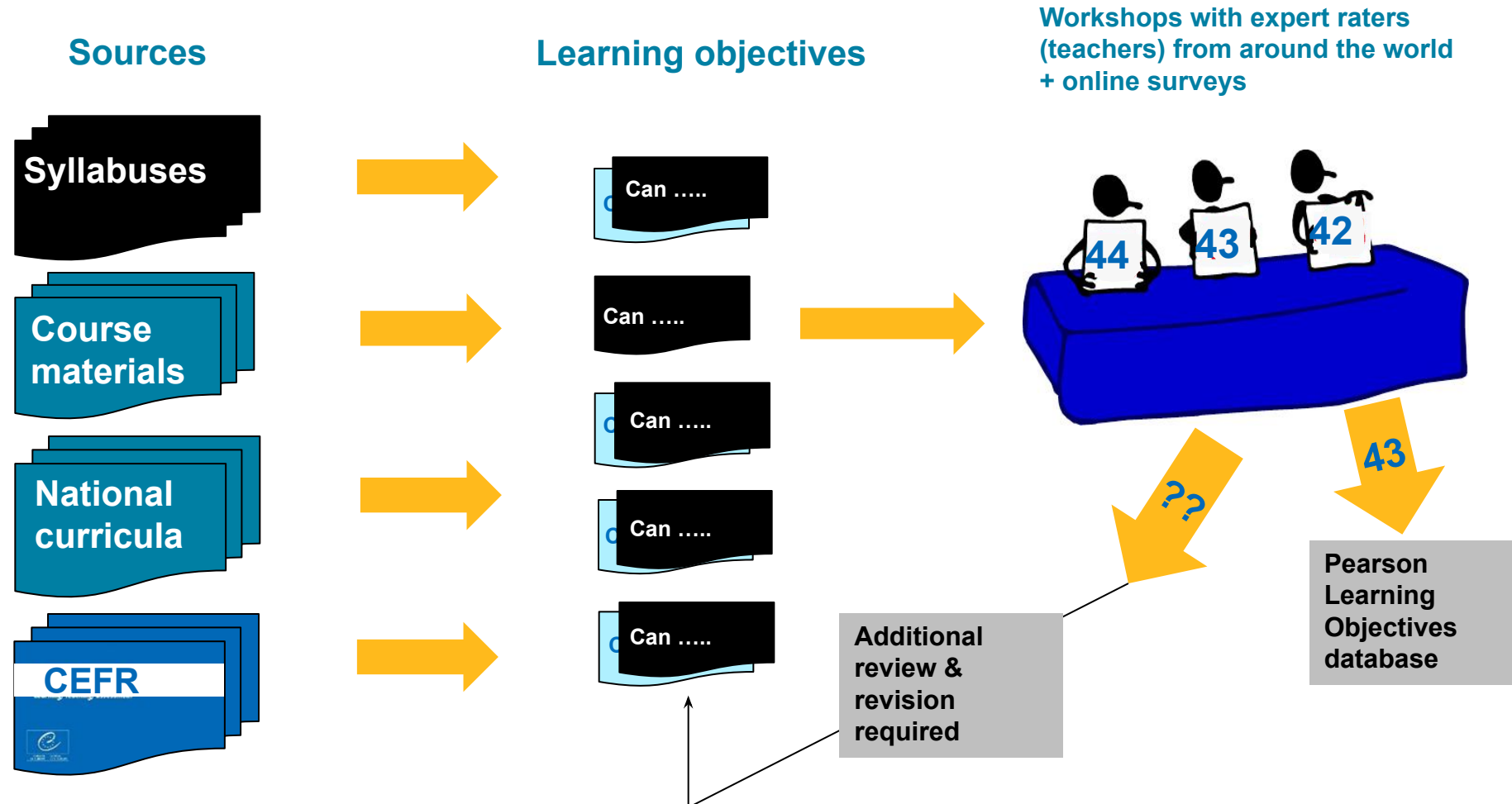
B1(+) [=GSE 51-58]

51



Creating Learning Objectives

Defining what a learner “can do” at specific levels of the Global Scale of English



1 A Work in pairs. Look at photos A-E. Which of the things in the box can you see?

- a mountain a hill a river a lake
a forest a village a market
a famous building a desert



- B** Work in pairs. Look at the words in the box in Exercise 1A and write an example for each word. Use your country if possible.

A mountain: Monte Cervino (Italy)

The Silk Route Bike Tour: is it the greatest journey in the world? The company calls it 'the longest, hardest, highest, hottest, coldest bicycle tour in the world'. In eighteen weeks you travel over 12,000 kilometres through seven different countries and have the most exciting ride of your life.

At night, you may find ancient roadside beds or in camp. A chef cooks dinner for about half of the 133 days; on other days cyclists can eat in local restaurants or shop in local markets for their food.

One cyclist blinged, 'The best thing was the friendliness of the people.' And the second

From Shanghai, you go across Asia, through Iran and arrive in Istanbul, Turkey, at the end. You travel across the deserts of China and cycle the Pamir highway to over 4,600 metres, across the 'Roof of the World' in the mountains of Tajikistan. On the way there are villages, markets and ancient buildings.

At night, you may in ancient
medieval beds or in camp.
A chef cooks dinner for about
half of the 1.33 days; on other
days cyclists can eat in local
restaurants or sleep in local
quarters like the hotel.

One cyclist blogged, 'The best thing was the friendliness of the people.' And the worst things? 'The snow and the cold in the Darnley.'

So is it the greatest journey in the world? Cyclist and blogger Miles MacDonald thinks so. 'The Silk Route is a true expedition. It is long and difficult, but it's such an amazing experience that you arrive in Istanbul a completely changed person.'

READING

- 2 A** Read the introduction to the article about the Silk Route Tour and answer the questions.

- 1 What is the Silk Route Tour?
- 2 How many countries does it travel through?

- B** Work in pairs. What else would you like to know about the Silk Route Tour? Write three questions using the words in the box to help.

countries	sleep	people	food	billus	price
-----------	-------	--------	------	--------	-------

Which countries does it visit?

- 3 A** Read the article. Did you find the answers to your questions?
- B** Read the article again. Are sentences 1–6 true (T) or false (F)?
- 1 The Silk Route Tour travels through seven countries in eighty weeks.
 - 2 A bicycle is faster than a camel.
 - 3 The tour starts in Istanbul and ends in Shanghai.
 - 4 Cyclists stay in hotels and camps.
 - 5 They take turns cooking dinner every evening.
 - 6 One cycle turns the snow and the hills were the best things.
- C** Would you like to go on the Silk Route Tour? Why/Why not?

The Silk Route Bike Tour



GRAMMAR

- 4 A** Complete the sentences with words from the article above.
- 1 Is it the _____ journey in the world?
 - 2 You see some of the _____ places in the world.
 - 3 The _____ thing was the friendliness of the people.

- B** Underline other examples of superlatives in the introduction. Then complete the table below.

	Adjective	Superlative	Rule
short adjectives	great	<u>the greatest</u>	Use _____ + adjective
longer adjectives	exciting	_____	Use _____ + adjective
irregular adjectives	good bad	_____ _____	

- C E 7.4 WEAK FORM:** the Listen to the pronunciation of the in the sentences in Exercise 4A. Then listen and repeat.

▶ page 140 **LANGUAGEBANK**

5 A Make questions about the Silk Route Tour.

- 1 What / cold / place you visited?
What was the coldest place you visited?
- 2 What / hot / place?
- 3 What / friendly / place?
- 4 What / far / distance you travelled in one day?
- 5 What / beautiful / building you saw?
- 6 What / amazing / experience of the journey?

- B** Match answers a)–f) with questions 1–6.

- 'It was Turpan, China'
'For me, the Gobi Desert: fifty degrees centigrade!'
- 'The Blue Mosque: I took 300 photos.'
'The Blue Mosque, Istanbul.'
- 'Meeting locals everywhere.'
'Arriving in Istanbul.'
- 'About 150 kilometres, in western China.'
'I don't remember ... at least 100 kilometres.'
- 'The mountains in Kyrgyzstan.'
'The desert: at night, there was snow!'
- 'Iran; the people there were so kind.'
'Everywhere, really.'

- C** **7.5** Listen to a conversation with a Silk Route Tour cyclist. Underline the answers he gives in a)–f).

SPEAKING

- 6 A** Work with other students. Plan a long journey to another country. Choose five places to visit: the most exciting, the most beautiful, the highest, etc.
 If we go to ..., then we visit the oldest/most famous ...

- 2** **B** Tell the class about your journey. Ask and answer questions about each journey.
A: Where do you sleep at night?
B: in small hotels.
C Discuss. Which journey is the most interesting?

WRITING

CHECKING AND CORRECTING

- 7 A** Read the student's homework below. Find and correct ten mistakes with:
- spelling • past simple • singular/plural
 - comparatives/superlatives

On Saturday we went by bus across the ^{country} ~~land~~.
We meet a lot of people. The people in the villages were
friendlier than in the city. At night we staid in a camp.
It was not very comfortable, but it was more cheaper
than the hotels. We bought all our food in markets.

- B** Write four sentences about your last holiday.

Aligning courseware

7.2 THE LONGEST BIKE RIDE

VOCABULARY

PLACES

1 A Work in pairs. Look at photos A-E.

B Work in pairs. Look at the words in the box in Exercise 1A and write an example for each word. Use your country if possible.

A mountain: Monte Cervino (Italy)

The Silk Route Bike Tour: is it the greatest journey in the world? The company calls it 'the longest, hardest, highest, hottest, coldest bicycle tour in the world'. In eighteen weeks you travel over 12,000 kilometres through seven different countries and have the most...

READING

2 A Read the Introduction to the article about the Silk Route Tour and answer the questions.

1 What is the Silk Route Tour?
2 How many countries does it travel through?

B Work in pairs. Use the box to help.

countries: slow, hot, cold, friendly, plain, far, beautiful, amazing, experience, snow, cold, hard, long, difficult, amazing experience, arrive in Istanbul, a completely changed person.

Which countries...

3 A Read the article.

B Read the article.

1 The Silk Route Tour is the longest journey in the world.
2 A bicycle is faster than a camel.
3 The tour starts in Istanbul and ends in Shanghai.
4 Cyclists stay in hotels and camps.
5 They take turns cooking dinner every evening.
6 One cyclist thought the snow and the hills were the best things.

C Would you like to go on the Silk Route Tour? Why/Why not?

The Silk Route Bike Tour

GRAMMAR

SUPERLATIVES

4 A Complete the sentences with words from the box.

1 Is it the _____ journey in the world?
2 You see some of the _____ places in the world.

of the people.

atives in the below.

Rule
Use _____ adjective
Use _____ adjective
Use _____ adjective

C 7 A WEAK FORMS: Listen to the pronunciation of the words in the box. Exercise 4A. Then listen and repeat.

page 340 LANGUAGEBANK

5 A Make questions about the Silk Route Tour.

1 What / cold / place you visited?
What was the coldest place you visited?
2 What / hot / place?
3 What / friendly / place?
4 What / far / distance you travelled in one day?
5 What / beautiful / building you saw?
6 What / amazing / experience of the journey?

B Match answers a)-f) with questions 1-6.

a) It was Turpan, China!
For me, the Gobi Desert: fifty degrees centigrade!

Work with other students. Plan a long journey to another country. Choose five places to visit: the most exciting, the most beautiful, the highest, etc.

First we go to _____, then we visit the oldest/most famous _____ for...

2 B Tell the class about your journey. Ask and answer questions about each journey.

A: Where do you sleep at night?
B: In small hotels.

C Discuss. Which journey is the most interesting?

WRITING

CHECKING AND CORRECTING

7 A Read the student's homework below. Find and correct ten mistakes with:

• spelling • past simple • singular/plural • comparatives/superlatives

On Saturday we went by bus across the desert. We meet a lot of people. The people in the village was friendlier than in the city. At night we stay in a camp. It was not very comfortable, but it was more cheaper than the hotels. We bought all our food in markets.

3 B Write four sentences about your last holiday.

Can understand the general meaning of short simple informational material and descriptions if there is visual support. (34)

Can ask and answer questions about basic plans and intentions. (38)

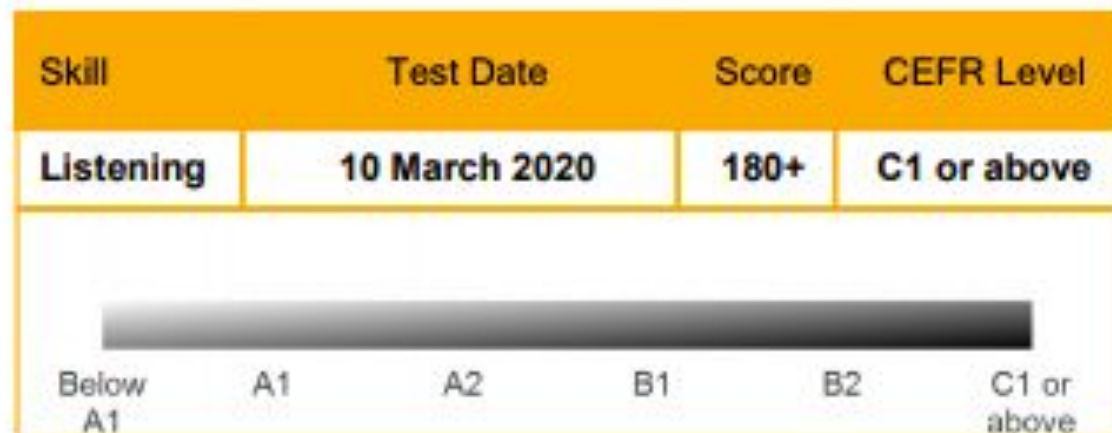
Can write short basic descriptions of past events and activities. (39)



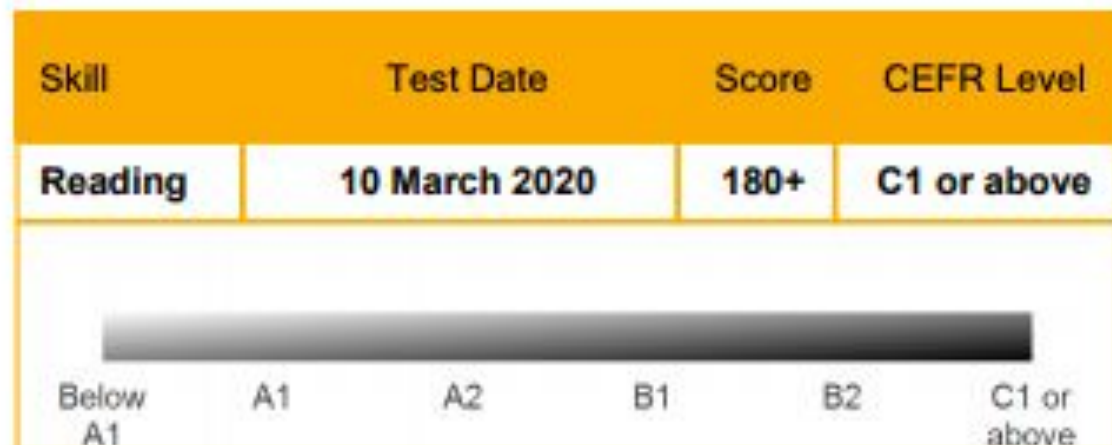
The real deal!

Assessment in the Classroom





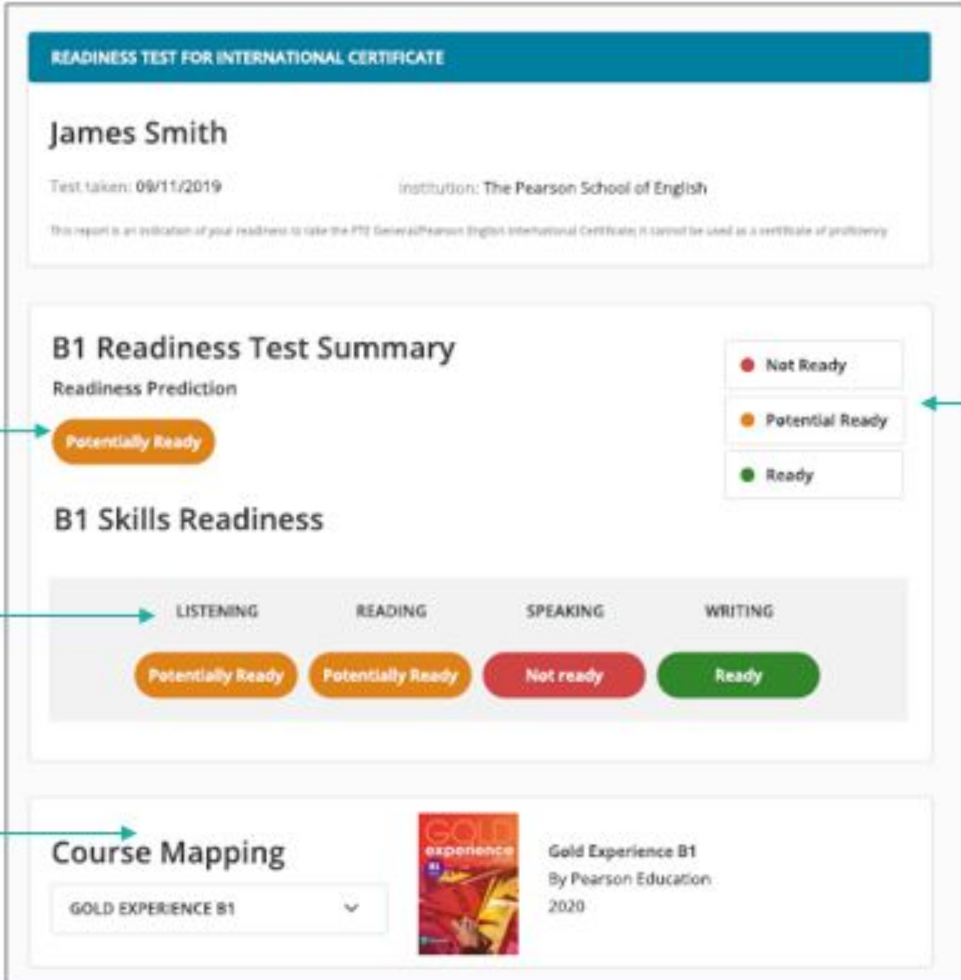
Can follow complex spoken language related to daily life and work and unfamiliar topics. Can extract details and key information, and infer intentions that are not explicitly stated. Can follow the sense of spoken information even when it is not clearly structured.



Can understand long and complex texts on a wide range of topics in daily life and work, including unfamiliar and abstract. Can extract key information and details, and comprehend information that is implied. Can detect the writer's tone and point of view.

Individual Readiness Report

In just a few short minutes, receive clear, detailed insights into your students' exam readiness and how you can help them reach the level they need to pass, across all four language skills.



The screenshot displays the 'Individual Readiness Report' for James Smith. The report is titled 'READINESS TEST FOR INTERNATIONAL CERTIFICATE' and includes the student's name, test date (09/11/2019), and institution (The Pearson School of English). A disclaimer states: 'This report is an indication of your readiness to take the PTQ General Pearson English International Certificate; it cannot be used as a certificate of proficiency.'

The 'B1 Readiness Test Summary' section shows an overall 'Readiness Prediction' of 'Potentially Ready' (orange button). To the right, a legend indicates three readiness levels: 'Not Ready' (red dot), 'Potential Ready' (orange dot), and 'Ready' (green dot).

The 'B1 Skills Readiness' section shows readiness for four skills: LISTENING (Potentially Ready), READING (Potentially Ready), SPEAKING (Not ready), and WRITING (Ready).

The 'Course Mapping' section includes a dropdown menu set to 'GOLD EXPERIENCE B1' and a book cover for 'Gold Experience B1 By Pearson Education 2020'.

Annotations with arrows point to specific features:

- 'Overall readiness indicator' points to the 'Potentially Ready' button in the B1 Readiness Test Summary.
- 'Readiness indicators for each language skill' points to the skill-specific readiness buttons in the B1 Skills Readiness section.
- 'If you're using a Pearson English coursebook, select it here' points to the 'GOLD EXPERIENCE B1' dropdown menu.
- 'Clear, at-a-glance 'traffic light' readiness indicators' points to the legend in the B1 Readiness Test Summary.

This is sample data for illustrative purposes only.

Individual Readiness Report cont.

In just a few short minutes, receive clear, detailed insights into your students' exam readiness and how you can help them reach the level they need to pass, across all four language skills.

Performance summary for each language skill

Listening

PERFORMANCE SUMMARY **Potentially Ready**

At this level, you can understand narratives and conversations about familiar topics, and can grasp some points of less familiar topics. You can follow most of an everyday conversation spoken at a natural pace if the speakers avoid very idiomatic language. You can follow changes of topic in discussions related to your fields of interest, and can recognise a speaker's feelings or attitudes.

RECOMMENDED ACTIVITIES FOR **B1**

Your listening skills are potentially at the level. Before taking the test, you should:

- Practice listening to online talks, lectures and other recordings until you are able to follow rapid, extended speech and distinguish between different points in a discussion or presentation.
- Develop the ability to work out the meaning of unknown words, topics and concepts by thinking about their context, considering what speakers have already said and what they go on to talk about.
- Listen to a variety of speakers to build familiarity with different accents and styles of speech and to develop the ability to infer speakers' opinions.

EXERCISES

- Unit 5 Listening Power up 3 • Unit 2 Listening Listen up 2 • Unit 7 Listening power up 2
- Unit 9 Speak up 5-7 • Unit 9 Listening Power up 2 • Unit 2 Power up 3-5 • Unit 7 Speaking Power up 2 • Unit 9 Listening Power up 4-5 • Unit 6 Listening Listen up 2-4
- Unit 2 Grammar 2 • Unit 2 Grammar 3 • Unit 1 Listen up 3

Reading

PERFORMANCE SUMMARY **Ready**

At this level, you can follow chronological sequence in a formal structured text. You can usually understand details of events, feelings and wishes in letters, emails and online postings. You can distinguish between fact and opinion, infer meaning based on information in a text, recognise examples and their relation to the main idea they support.

Performance indicator for each language skill

Recommended activities targeted to learners' weaker areas, compared to the level required in their exam

Specific exercises in your chosen coursebook that will help learners target the skills gaps identified in the test

This is sample data for illustrative purposes only.

Group Readiness Report

In just a few short minutes, receive clear, detailed insights into your students' exam readiness and how you can help them reach the level they need to pass, across all four language skills.

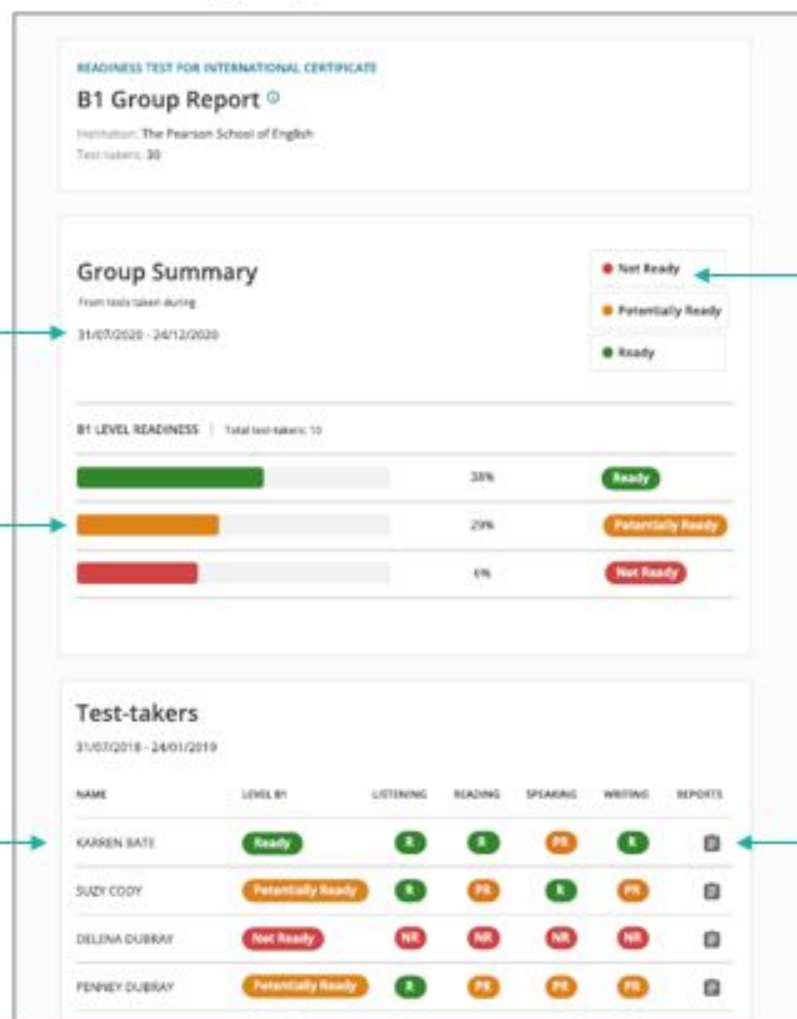
Testing period

Overview of your whole group's exam readiness with percentages given for each status

Readiness statuses of each individual in the group clearly listed for easy comparison

Same clear traffic light indicators for each stage of readiness

Links through to individual student reports



This is sample data for illustrative purposes only.

Group Readiness Report cont.

In just a few short minutes, receive clear, detailed insights into your students' exam readiness and how you can help them reach the level they need to pass, across all four language skills.

Breakdown of your class' readiness in reading, writing, speaking and listening, with an average performance summary for each skill

Recommended activities targeted to the weaker areas identified in your class as a whole

If you're using a Pearson English coursebook with your class, select it here

Overall readiness indicator per skill

Specific recommended exercises in your choice of Pearson coursebook that will directly target skills gaps in your class



Course Mapping

GOLD EXPERIENCE B1

Gold Experience B1
By Pearson Education
2020

Listening

PERFORMANCE SUMMARY **Ready**

Test takers at this level are generally able to understand narratives and conversations about familiar topics, and can grasp some points of less familiar topics. They can follow most of an everyday conversation spoken at a natural pace if the speakers avoid very idiomatic language. They can generally follow changes of topic in discussions related to their fields of interest, and can recognise a speaker's feelings or attitudes.

RECOMMENDED ACTIVITIES FOR B1

The test takers' listening skills are potentially at the level. Before taking the test they should:

- Practice listening to online talks, lectures and other recordings until they are able to follow rapid, extended speech and distinguish between different points in a discussion or presentation.
- Develop the ability to work out the meaning of unknown words, topics and concepts by thinking about their context, considering what speakers have already said and what they go on to talk about.
- Listen to a variety of speakers to build familiarity with different accents and styles of speech and to develop the ability to infer speakers' opinions.

EXERCISES

• Unit 5 Listening Power up 3 • Unit 2 Listening Listen up 2 • Unit 7 Listening Power up 2 • Unit 9 Speak up 5-7 • Unit 9 Listening Power up 2 • Unit 2 Power up 3-5 • Unit 7 Speaking Power up 2 • Unit 9 Listening Power up 4-5 • Unit 6 Listening Listen up 2-4 • Unit 2 Grammar 2 • Unit 2 Grammar 3 • Unit 1 Listen up 3

Reading

PERFORMANCE SUMMARY **Ready**

Test takers at this level can generally follow chronological sequence in a formal structured text. They are

This is sample data for illustrative purposes only.

To sum up

- Address construct under-representation
- Context driven assessment
- More meaningful and accurate scoring based on integrated skills
- Faster and more accurate, more detailed and varied feedback
- Freeing the teacher to do what they do best

Questions?





Pearson