

Context Matters

Choosing the Right Language Test



by  Paragon

Introductory Discussion

Discrete-point Tests
Task-based Assessment
Integrated Task-based Assessment

Different Types of Language Tests

Discrete-point Tests

- Assumes language can be reduced to separate elements that can be individually assessed.
- Learning involves acquisition of a finite set of discrete elements.
- Items intend to test one language element at a time by isolating it from the broader context.

Task-based Assessment

- Views language as highly contextualized & goal-driven.
- Captures and assesses language *in use*.
- Measures complex abilities beyond a single trait.

Independent Tasks

- Tasks focus on one component of language (S,W,L,R), but involve the use of many skills within that component.



Integrated Tasks

- Tasks integrate two or more language components.
- Involves input → output
R/L → W/S



Examples

Task-based Assessment

Discrete-point Tests

Sentence completion targeting specific grammar points.

e.g. Climate change_____ (BE)
an important topic these days.

Independent Task

Writing task on a provided topic.

e.g. *In your opinion*, do you think the problem of climate change can ever be solved?

Integrated Task

Writing task based on provided source material.

e.g. *According to the passage*, what are the most critical factors in climate change?



Discussion

1. How would you describe the following types of language tests?
 - Discrete point tests
 - Task-based assessments
 - Integrated task-based assessments

2. Discuss your experience working with these different types of tests.
 - Discrete point tests
 - Task-based assessments
 - Integrated task-based assessments

Discrete-point Testing

Isolates language elements to be measured separately

Use

- measure knowledge & ability with specific language points

Design

- items isolate specific elements of language
- items tend to be decontextualized
- criteria for success = *correct selection or production of the target language point*

Discrete-point Testing

Isolates language elements to be measured separately

Advantages

- measure a broad range of target language points
- practical to administer
- easy to score
- provide fine-grained information

Drawbacks

- appear easy to construct, but difficult to construct well (Purpura, 2004)
- results may not generalize to language in use

Types of Discrete-point Items

Selected Response Items

- Choose from two or more options
 - Noticing – circle the nouns
 - Matching
 - True/False
 - Multiple choice
 - Categorize or group
- Criteria for success – the ability to *identify* the correct response
- Test scores are informative about the candidate's *receptive* knowledge

Constructed Response Items

- Limited production
 - Label
 - Cloze
 - Sentence Completion
- Criteria for success – the ability to *produce* a correct response
- Test scores are informative about the candidate's language *production*

Task-based Assessment

Measures language in use; ability to use language to achieve a goal

Use

- measure ability to perform a communicative activity

Design

- contextualized; placed within a particular language use setting
- aim to simulate the real-life target domain
- assess receptive or productive skills
- criterion for success = *theoretically*, the ability to perform the task
 - But, *in reality*, task performance is often measured by specific criteria

Integrated Task-based Assessment

Measures ability to integrate language skills to achieve a goal

Use

- measure ability to perform a communicative activity *involving 2+ language skills*

Design

- tasks are contextualized; simulate the real-life target domain
- activities involve the integration of receptive & productive skills
 - e.g. Reading/Listening → Speaking/Writing
- criterion for success = *theoretically*, the ability to perform the task

Productive Task-based Assessment

Independent vs Integrated S/W Tasks
Content Development & Scoring Perspectives

Independent Speaking or Writing Tasks

Tasks that function on their own

Design

- ask general questions or present familiar situations
- answer based on life experience, common knowledge, or personal opinions

Advantages

- Ease relatively easy to develop
- Familiarity tasks feel natural for test takers
- Positive washback encourage diverse teaching & learning methods
- Authenticity provide realistic context for language use

Independent Speaking or Writing Tasks

Tasks that function on their own

Drawbacks

- assume some level of background knowledge or experience with the topic
- require some degree of creativity on the part of the test taker
- restricted in scope; limited to general topics
 - reduction in task authenticity (especially in academic context)
 - under-representation of the domain of interest

Integrated Speaking or Writing Tasks

Tasks that incorporate source materials

Design

- questions related to content of source materials (reading or listening)
- answers based on information provided in sources (content & language input)

Advantages

- **Fairness** mitigate disadvantages from lack of topic knowledge
 - *reduces topic anxiety & creative demands*
- Positive washback encourage diverse teaching & learning methods
- Authenticity provide realistic context for language use

Integrated Speaking or Writing Tasks

Tasks that incorporate source materials

Drawbacks

- more challenging to develop
- more challenging to rate
 - may complicate measurement of productive & receptive skills
- issues with copied content from source materials
- may measure different skills than independent tasks

Considerations in Content Development

1. Consider the influence of the *task* on S/W performance.
 - With all tasks, the complexity of the prompt can impact test taker performance.
 - Test developers should design prompts to be clear and accessible.
 - Consider:
 - difficulty of language in the prompt (how the question/instructions are worded)
 - complexity of the skills required (what TTs need to do to complete the task)

Considerations in Content Development

2. Consider the influence of the *source materials* on S/W performance.

- Complexity & density of the R/L sources can impact test taker performance.
 - reading plays a role in R→W tasks (Plakans 2009)
- Test developers have to choose R/L sources carefully.
 - determine appropriate content for S/W task
- Consider:
 - length (length of the passage/target section)
 - complexity (clarity, sophistication of the language)
 - density (amount of information/number of ideas)

Considerations in Content Development

3. Keep the objective of the integrated task in mind.

- Integrated (R/L → S/W) tasks are generally considered *speaking* or *writing* tasks.
 - primary purpose is *not* to re-assess reading/listening comprehension
 - aim to mitigate the impact of source materials on TTs performance
- Consider the purpose of the questions.
 - primary purpose is to elicit a S/W performance
 - ask questions that facilitate speaking & writing production

Considerations in Content Development

4. Select source material content that is *accessible*. (Gebriel & Plakans 2009)

- For reading-based tasks:
 - focus on main ideas from the passage
 - avoid highly complex or technical content
- For listening-based tasks:
 - focus on salient/memorable information from the recording
 - avoid very specific or trivial details

Considerations in Scoring

Speaking and Writing Tasks

- Speaking or writing performance is the focus of the assessment.
- Evaluation must take into account the multiple skills engaged but primarily evaluate Speaking or Writing proficiency.
- Score report of Listening, Reading, Writing, and Speaking proficiency

Considerations in Scoring

1. Comprehension of the input - the prompt and/or source material - can impact the response.

- Does the response address the prompt?
 - The response must be in some way related to prompt.
 - Raters are open to varied interpretations of the prompt (within reason).
 - This is one element of the response that is assessed.
- Does the response demonstrate use or understanding of the source material?
 - This is one element of the response that is assessed.
 - However, meaning may be communicated and language use demonstrated in the response despite misunderstanding, misuse, or no use of source material.

Considerations in Scoring

2. A Speaking or Writing performance is more than an answer to a question.

- Achieving the communicative purpose of the task is important and is evaluated.
- Rating a performance involves more than deciding if an answer is right or wrong.
- Although raters consider task completion and source use, they evaluate meaning and language use in test taker performances.

Considerations in Scoring

3. Raters must deal with the use of sources.

- Distinguish Test Taker production from that borrowed from Source Material
 - Language use borrowed from source material must be identified.
 - Consider appropriate re-use of key terminology but recognize over-reliance on source lexicon or grammatical structures.
 - Evaluate the language use of the test taker.

Considerations in Scoring

4. Evaluate source use from the perspective of proficiency rather than plagiarism.

- In the assessment context, verbatim copying in a response may reflect a stage in developing proficiency of source use.
 - Lower proficiency responses – may have extensive verbatim copying
 - consider selection of relevant vs irrelevant ideas
 - minimal modification – word swapping
 - Mid to high proficiency responses – modification and integration
 - paraphrasing and/or summarizing
 - integration or synthesis of ideas from sources

Research Findings

Integrated Speaking Tasks
Integrated Writing Tasks

Research Findings on Integrated *Speaking* Tasks

1. Integrated speaking tasks effectively distinguish speaking proficiency. (Frost et al 2011)
2. Performance on integrated speaking tasks: (Brown et al 2005)
 - more complex rhetorical structures & higher quality of ideas
 - no differences in grammar & vocabulary
 - quality of response depended on complexity of source material
 - responses show more complex functional & organizational skills
3. Test takers use a wider range of strategies on integrated tasks. (Barkaoui et al 2012)
 - suggests integrated tasks require broader range of language skills
 - better measure the '*strategic competence*' aspect of communicative ability

Research Findings on Integrated *Speaking* Tasks

4. Scores on independent & integrated speaking tasks correlate strongly. (Lee 2006)
5. Research supports the use of both independent & integrated tasks. (Barkaoui et al 2012)
 - Choice of task depends on the objective of the assessment:
 - what language skills & processes the test intends to measure
 - the demands of the target real-life language context

Research Findings on Integrated *Writing* Tasks

1. Use of source materials varies by proficiency. (Gebril and Plakans, 2009)
 - summary & better paraphrasing more common with high proficiency
 - 'attempted paraphrase' common with lower proficiency
 - lowest proficiency test takers may not use sources at all
2. Verbatim copying of sources is related to proficiency. (Gebril and Plakans, 2009)
 - verbatim source use tends to occur with lower proficiency writers

Research Findings on Integrated *Writing* Tasks

3. Processes used to respond to integrated prompts vary by proficiency (Plakans, 2009)
- high proficiency writers use more strategies than low proficiency writers

“There are *two aspects of test performance* that we need to investigate in our evaluation of test usefulness: *the processes* or strategies that test takers use in responding to specific test tasks and *the product* of those processes or strategies. In order to evaluate the usefulness of a given test, we need to investigate both aspects.” (Bachman, 2002, p.5)

Final Thoughts

Integration as a Spectrum
Choosing the 'Right' Test
The Test Validity Perspective

Our Perspective

- Most tasks are integrated *to some extent* – i.e. involve more than one language skill
 - Even ‘independent’ S/W tasks rely somewhat on reading skills.
- Tasks are often categorized as being ‘independent’ or ‘integrated’, but...
 - It may be more useful to think about integration *as a spectrum*.
- Teachers/assessors can incorporate varying degrees of integration into language tasks, depending on the purpose of the assessment.

Choosing the Right Language Test

- There is not one 'right' test.
- Different types of tests are designed to suit different situations & purposes.
- The choice of test depends on:
 - the practical constraints *the test location, time, delivery mode & available resources*
 - the objective of the assessment *the language skills/processes the test is trying to measure*
 - the domain of interest *the real-life context/demands the test intends to simulate*
 - the information required *the claims the assessor wants to make about test takers*

The Test Validity Perspective

- Language tests cannot capture all aspects of test taker ability.
 - language ability is measured based on evidence (test performance)
- This involves making inferences about test takers' true abilities.
 - score on the test → claims about real-life ability



The Test Validity Perspective

- Task-based language assessment (TBLA) simulates real-life language use (Ellis, 2003)
 - TBLA captures language *in use* – to achieve a goal in a particular context
- Context influences all aspects of real-life language use
 - Production: language choices / Reception: way that meaning is interpreted
- TBLA helps close the gap between test performance & real-life ability
 - Strengthens inferences about test taker ability based on test scores



The Test Validity Perspective

General Language Proficiency Assessment



Task-based
Assessment

Short answer task:

'You missed work yesterday because you were sick. Explain to your boss why you were absent.'



Real-life speaking or writing ability

Discrete
Point Test

Sentence completion:

e.g. I _____ (miss) work yesterday because I was sick.



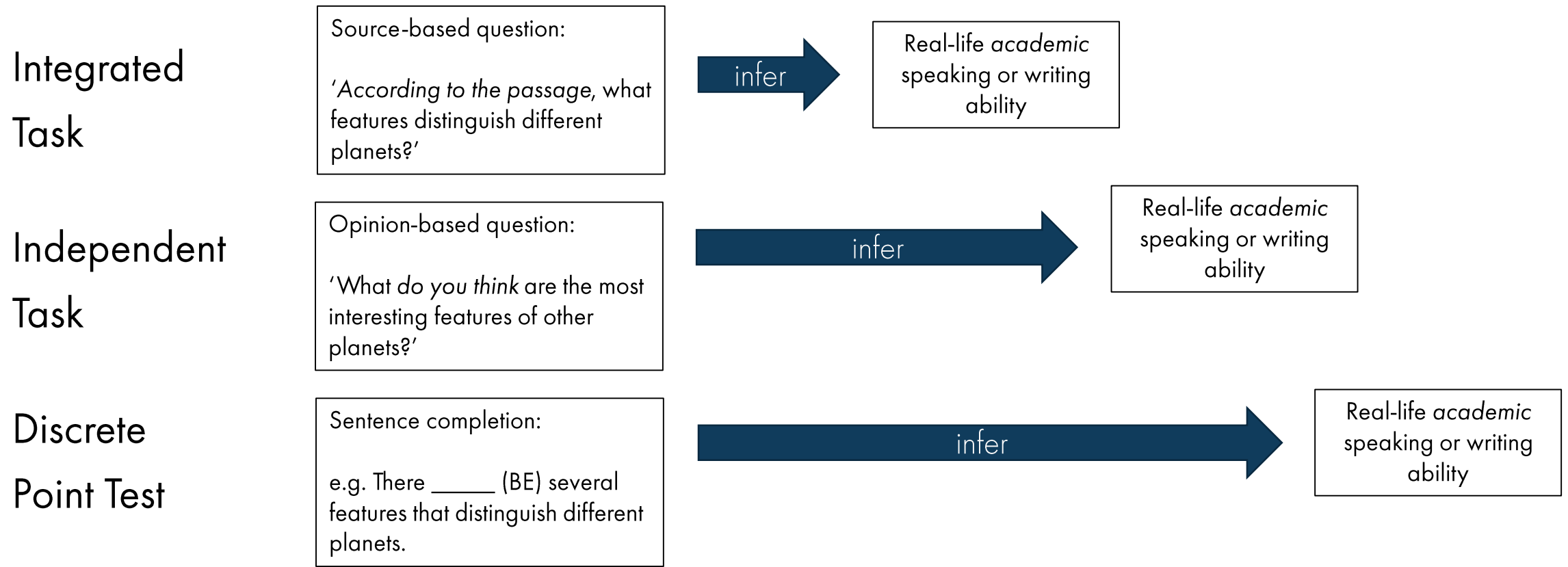
Real-life speaking or writing ability



CAEL[®]

The Test Validity Perspective

Academic Language Proficiency Assessment



Considerations for Task-based Assessment

Points to consider when selecting/designing tasks:

1. Does the task engage the *areas of language* we want to assess?
2. Does completing the task require the same *cognitive processes* as real life?
3. Does the task provide *adequate context* to assess language in use?
4. Does the task reflect the same *level of demand* as real life?

Additional considerations:

1. Is the task *fair and appropriate* for the test takers?
2. Does the task require *prior knowledge or experience* that all TT might not possess?
 - *Integrated tasks are one way to mitigate this!*

Integrated Tasks in the Academic Context

- Integrated tasks more closely reflect real-life academic tasks (Hinkel, 2006)
 - students speak & write about class content (textbooks, lectures)
 - language skills are engaged in an integrated way
- Integrated tasks support stronger inferences about real-life academic ability
 - cognitive processes required for task reflect real-life processes
 - task demands match that of real-life academic contexts
- Integrated tasks mitigate fairness concerns w/academic content in language testing
 - cannot assume test taker knowledge/experience with academic subject matter
 - source materials provide a level playing field for all test takers

Follow-up Discussion

Advantages & Challenges of Different Tests
Considerations in Choosing a Test

Final Discussion

1. What are some of the advantages and/or challenges associated with the following types of tests from your perspective?

- Discrete point tests
- Task-based assessments
- Integrated assessments

2. What are the most important considerations in choosing a language test in your language teaching/assessment context?

Discrete-point Tests

Advantages

- Focus on Form drawing learners' attention to specific elements of language
- Scoring easier to score in a reliable manner

Challenges

- Development claimed to be easier to develop, but not always the case
- Authenticity may not adequately capture real-life language use
- Validity may not adequately assess communicative competence

Task-based Assessment

Advantages

- Authenticity language used to achieve a goal; real-life contexts & demands
- Engagement more interactive & engaging for test takers
- Validity more adequately measures communicative competence

Challenges

- Development requires detailed specifications (outline of features)
 - language features & processes, context, demand
- Scoring requires well-developed rubric
 - criteria for task completion

Integrated Task-based Assessment

Advantages

- Authenticity more representative of real-life demands - *especially academic*
- Engagement more interactive & engaging for test takers
- Validity more adequately measures academic language proficiency

Challenges

- Development must consider relationship between source materials & task
 - R/L sources impact S/W performance
- Scoring requires well-developed rubric – *especially for use of sources*

References

- Bachman, L. F. (2002). Some Reflections on Task-Based Language Performance Assessment, *Language Testing*, 19, p. 453.
- Barkaoui, K., Brooks, L., Swain, M., & Lapkin, S. (2012). Test-Takers' Strategic Behaviors in Independent and Integrated Speaking Tasks, *Applied Linguistics*, 1, 22.
- Brown, A., Iwashita, N., & McNamara, T. (2005). An Examination of Rater Orientations and Test Taker Performances on EAP Speaking Tasks, *Monograph Series*, 29, Educational Testing Service.
- Ellis, R. (2003). Designing a Task-Based Syllabus, *RELC Journal*, 34, 1, p.64-81.
- Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the Validity of an Integrated Listening-Speaking Task, *Language Testing*, 29, 3, p. 345–369.
- Gebril, A. & Plakans, L. (2009). Investigating Source Use, Discourse Features, and Process in Integrated Writing, *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 7, p. 47-84.
- Hinkel, E. (2006). Current perspectives on teaching the four skills, *TESOL Quarterly*, 40, 1, p.109-131.
- Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks, *Language Testing*, 23, p. 131-166.
- Plakans, L. (2009). The Role of Reading Strategies in Integrated L2 Writing Tasks, *Journal of English for Academic Purposes*, 8, 4, p.252– 266.
- Purpura, J. (2004). *Assessing Grammar*. Cambridge, U.K.: Cambridge University Press.

Contact

Katharine Wallace, Scoring Lead
Jennifer Flasko, Content Development Lead

Visit: paragontesting.ca
Email: research@paragontesting.ca