How do we assess task-based performance? Recent advances in language assessment

LARC/CALPER webinar

April 3, 2014

John M. Norris - Georgetown University

Copyright © John M. Norris, 2014: Please cite as:

Plan:

0. What is assessment and why do we do it?
1. Traditions, alternatives → Task-based assessment
2. The range of uses for task-based language assessment
3. Issues in assessing task-based language performance
4. Some challenges and opportunities in TBLA

Task-based language assessment: the elicitation and evaluation of language use (across all modalities) for expressing and interpreting meaning, within a well-defined communicative context (and audience), for a clear purpose, towards a valued goal or outcome.
What is assessment?

What isn’t assessment (in language education)?

CEFR
B1
What is assessment and why do we do it?

What is educational assessment?


Assessment is itself a particular kind of narrative:

- an evidentiary argument
- about aspects of what students know and can do,
- based on a handful of particular things they have said, done, or made.

Which narratives do we want to tell in language assessment?
What is assessment and why do it?

1. Fill in the missing letters to create complete words that make sense in the passage.

This is an example C-test passage. Starting with the second word of this sentence, the last half from each consecutive word has been deleted. C-tests are typically composed of multiple texts which become increasingly difficult. Each text usually contains between 20 and 25 words. So, which do you think is tested on a C-test?

Which is the ‘better’ language assessment?

2. The European Union and the United States: Possible Comparisons and Lessons

You have been invited to give a lecture at the Society for German-American Relations in Weimar next month. For your contribution, the organizers are looking for a presentation on the much-discussed topic “The European Union and the Founding of the United States - A Model?” Your talk is limited to 15 minutes, to be followed by discussion of your remarks.
The ‘correct’ answer:
IT DEPENDS ON THE INTENDED USE of the TEST!
The ‘correct’ answer: IT DEPENDS ON THE INTENDED USE of the TEST!

Specifying intended uses for assessment (Norris, 2000)
1. Fill in the missing letters to create complete words that make sense in the passage.

This is an example C-test passage. Starting with the second word of this sentence, the last half from each consecutive word has been deleted. C-tests typically consist of multiple texts which become increasingly difficult. Each text usually contains between 20 and 25 items. So, what do you think is tested on a C-test?

Intended use: quick estimate of language proficiency, e.g., for use in placing students into language courses...not task-based assessment.

2. The European Union and the United States: Possible Comparisons and Lessons

You have been invited to give a lecture at the Society for German-American Relations in Weimar next month. For your contribution, the organizers are looking for a presentation on the much-discussed topic “The European Union and the Founding of the United States - A Model?” Your talk is limited to 15 minutes, to be followed by discussion of your remarks.
What is assessment and why do it?

1. Fill in the missing letters to create complete words that make sense in the passage.

This is an example C-test passage. Starting with the second word of this sentence, the last half of each consecutive word has been deleted. C-tests are typically composed of multiple texts that become increasingly difficult. Each text usually contains between 20 and 25 items. So, what do you think are tested on a C-test?

Intended use: quick estimate of language proficiency, e.g., for use in placing students into language courses...not task-based assessment.

2. The European Union and the United States: Possible Comparisons and Lessons

You have been invited to give a lecture at the Society for German-American Relations in Weimar next month. For your contribution, the organizers are looking for a presentation on the much-discussed topic “The European Union and the Founding of the United States – A Model?” Your talk is limited to 15 minutes, to be followed by discussion of your remarks.

Intended use: extended, authentic, and integrated performance with the target language, e.g., for end-of-unit achievement testing and feedback...yes, task-based assessment.
EDUCATIONAL USES FOR ASSESSMENT

Making decisions about students

- admissions
- placement
- course credit
- grading/achievement
- advancement/retention
- fulfilling requirements
- certification of abilities
- awarding
- funding

ETC.

INDIVIDUAL DECISION-MAKING
EDUCATIONAL USES FOR ASSESSMENT

INDIVIDUAL DECISION-MAKING

Making decisions about students: admissions, placement, course credit, grading/achievement, advancement/retention, fulfilling requirements, certification of abilities, awarding funding, ETC.

CLASSROOM FEEDBACK

Informing classroom teaching and learning: diagnosing learner needs, understanding learning outcomes, motivating students, defining learning expectations, improving instruction, focusing students’ learning/study, ETC.

ETC.
EDUCATIONAL USES FOR ASSESSMENT

Making decisions about students

INDIVIDUAL DECISION-MAKING

Understanding learning outcomes

Diagnosing learner needs

Improving, ensuring, demonstrating quality

Improving teacher performance

Aligning, articulating program practices

Communicating beyond, about the program

Responding to program review

ETC.

PROGRAM EVALUATION

Improving, ensuring, demonstrating quality

Improving, ensuring, demonstrating quality

Clarifying program mission/needs

ETC.

Revising curriculum/pedagogy

Improving teacher performance

Aligning, articulating program practices

Communicating beyond, about the program

Responding to program review

ETC.

PROFESSIONAL DECISION-MAKING

Understanding learning outcomes

Diagnosing learner needs

Improving, ensuring, demonstrating quality

Improving teacher performance

Aligning, articulating program practices

Communicating beyond, about the program

Responding to program review

ETC.

PROGRAM EVALUATION

Improving, ensuring, demonstrating quality

Improving, ensuring, demonstrating quality

Clarifying program mission/needs

ETC.

Revising curriculum/pedagogy

Improving teacher performance

Aligning, articulating program practices

Communicating beyond, about the program

Responding to program review

ETC.

PROFESSIONAL DECISION-MAKING

Understanding learning outcomes

Diagnosing learner needs

Improving, ensuring, demonstrating quality

Improving teacher performance

Aligning, articulating program practices

Communicating beyond, about the program

Responding to program review

ETC.

PROGRAM EVALUATION

Improving, ensuring, demonstrating quality

Improving, ensuring, demonstrating quality

Clarifying program mission/needs

ETC.

Revising curriculum/pedagogy

Improving teacher performance

Aligning, articulating program practices

Communicating beyond, about the program

Responding to program review

ETC.
EDUCATIONAL USES FOR ASSESSMENT

Making decisions about students
- admissions
- placement
- course credit
- grading /achievement
- advancement/retention
- fulfilling requirements
- certification of abilities
- awarding funding
- ETC.

Improving, ensuring, demonstrating quality
- demonstrating student learning
- revising curriculum/pedagogy
- improving teacher performance
- aligning, articulating program practices
- communicating beyond, about the program
- responding to program review
- clarifying program mission/needs
- ETC.

Informing classroom teaching and learning
- understanding learning outcomes
- motivating students
- defining learning expectations
- improving instruction
- focusing students’ learning/study
- diagnosing learner needs

Program Evaluation
- aligning, articulating program practices
- communicating beyond, about the program
- responding to program review
- clarifying program mission/needs
- ETC.

Classroom Feedback
- improving instruction
- motivating students
- defining learning expectations
- understanding learning outcomes

INDIVIDUAL DECISION-MAKING

What different kinds of assessment are needed to meet these different uses? How might tasks play a role?
What are the possibilities for assessment?

METHODS

- MCT
- Matching
- Fill-in-the-blank
- Short answer
- Cloze/C-test
- Performance assessment
- Observations
- True-false
- Projects
- Portfolios
- Essays
- Matching
- Interviews
- Rubrics
- Self-assessment
- Peer-assessment
- Fill-in-the-blank
- Short answer
- Self-assessment
- Peer-assessment
Alternatives in assessment: The move to task
Early interest in ‘tasks’ for assessment

“If we want to find out how well a person can perform a task, we can put him to work at that task, and observe how well he does it and the quality and quantity of the product he turns out.”

Cureton (1950): “Validity”

e.g., in language testing:

*Foreign Service Institute Oral Interview* (1950s)

Using tasks to elicit a holistic estimate of functional language ability, from 0 (no ability) to 5 (educated native speaker ability)
Standardized testing and psychometrics

But, the primary tradition of assessment in the 20th century:
Large-scale standardized testing: minimal competency, accountability
→ emphasis on efficiency, technology (e.g., scoring), discrete-point
→ primary concern with psychometrics, norm-referencing
→ validity = narrowly focused on construct representation, reliability

e.g., in language testing:

*Test of English as a Foreign Language (1964)*

MCT: vocabulary, reading, listening, structure, grammar (emphasis on testing language knowledge)

*But see later iterations of TOEFL...*
Influence of traditional testing

When we ask teachers what an assessment looks like, what do they tell us?

- True-false
- Matching
- Multiple choice
- Fill-in the blank
- Exam
- Test
- Giving grades
- Accountability
- ETC.
Influence of traditional testing

When we ask teachers what an assessment looks like, what do they tell us?

- True-false
- Matching
- Multiple choice
- Fill-in the blank
- Exam
- Test
- Giving grades
- Accountability
- ETC.

Washback of testing traditions:
- Narrowing teaching and learning
- Misrepresenting assessment
- Disconnect between teaching-testing-learning
Challenges to ‘traditional’ testing

e.g., Michael Byram (1997): Need for change in intercultural assessment

“It is the simplification of competencies to what can be ‘objectively’ tested which has a detrimental effect: the learning of trivial facts, the reduction of subtle understanding to generalizations and stereotypes, the lack of attention to interaction and engagement because these are not tested.”
Reactions to ‘traditional’ testing

- Portfolios
- Performance-based assessments
- Essays
- CBTs
- Self-assessments
- Interviews
- Rating scales, rubrics

1980s-90s
- Authenticity
- Performance
- Can-do (vs. know)
- Classroom
- Washback
The move to task as assessment medium

Grant Wiggins (1998): Educative assessment

The primary assessment medium should be authentic tasks which have learners demonstrate what they can do with what they are learning; enables meaningful integration of teaching, learning, and assessment → assessment becomes educative (informative) rather than just an authoritarian mechanism for control.

The move to task-based language assessment

Geoff Brindley (1994): Task-centered assessment

“...the process of evaluating, in relation to a set of explicitly stated criteria, the quality of the communicative performances elicited from learners as part of goal-directed, meaning-focused language use requiring the integration of skills and knowledge”

→ aligned with CLT, TBLT movements in pedagogy
→ emphasis on authenticity of L2 use
→ validity = capturing ability for use in valued tasks
→ classroom applications obvious, but beyond?
Operationalizing task-based language assessment


“...genuinely task-based language assessment takes the task itself as the fundamental unit of analysis, motivating item selection, test instrument construction, and the rating of task performance”

1. Specify intended uses for the assessment: who, what, why, impact
2. Select and analyze key target tasks/features from needs analysis
3. Design tests and items: emphasize authentic performances
4. Determine real-world criteria for rating task performance qualities
5. Pilot-test and revise instruments and procedures (raters/rating)
6. Evaluate validity in terms of intended uses, and especially impact on teaching and learning.
Why task-based language assessment?

Task-based language assessment: the elicitation and evaluation of language use (across all modalities) for expressing and interpreting meaning, within a well-defined communicative context (and audience), for a clear purpose, towards a valued goal or outcome.

Tasks respond to diverse needs/uses in assessment

• Focus on what L2 users can actually do with the language
• Integrate language form and meaning into what gets assessed
• Provide useful feedback to teachers, learners, curriculum
• ‘Wash back’ on teaching and learning by making outcomes ‘real’
• Counter negative influence of traditional knowledge assessment
• Align testing with new, evolving language pedagogies
Up Next:
How do we assess task-based performance in language classes and programs?
Issues, examples, suggestions...

But First:
Any Questions?
Issues

1. Which tasks do we assess?
2. How do we elicit task-based performance?
3. What gets assessed in task performance?
4. How do we score task performance?
1. Which tasks do we assess?

1. Write a recommendation request e-mail to professor
2. Write an e-mail to potential employer to send application packet
3. Write an e-mail to refuse professor’s request
4. Write constructive comments on cover letter written by classmate
5. Give oral peer-feedback on classmate’s request e-mail to professor
6. Interacting with classmates in situations of making suggestions and disagreement
7. Interacting with a professor in situations of making requests and refusal

Why these tasks?

e.g.: EAP pragmatics assessment (Youn, 2009)
1. Which tasks do we assess?

Which tasks?

→ Based on needs analysis for specific English language program context (individual & interactive tasks)
→ High priority tasks reflecting pragmatic challenges for college-level international students
→ Each task serves as end-of-unit achievement assessment
→ Combined, all of the tasks can be used for diagnostic assessment
You have been given the opportunity of a lifetime to apply for an athletic training camp in [e.g., Spain], tuition free! This camp trains young people in all sports, from the extreme (snowboarding, bicycle motor cross, rollerblading) to team sports of all kinds (basketball to volleyball). You name it, they help you train for it! To be accepted into the camp, all applicants have to convince the admissions office that they have good exercise and nutrition habits. First, you will read about health and nutrition from the perspective of the Spanish - speaking world. Then you will discuss your eating and exercise regimen with your partner to compare your nutrition and exercise, perhaps event to get some ideas. You will then write your application letter to the summer camp describing your nutrition and training regimen, convincing them that you are well prepared for the camp and need to be accepted.

Overall introduction to a sequence of interrelated pedagogic/assessment tasks. Note the distinct task phases.
1. Which tasks do we assess?

- Individual reading task
- Pair discussion task
- Letter writing task

Why these tasks?

Consortium for assessing performance standards
http://flenj.org/CAPS/?page=parent

e.g., Integrated performance assessment in US schools
1. Which tasks do we assess?

- **Individual reading task**
- **Pair discussion task**
- **Letter writing task**
- **Self-assessment:** understanding task content, language
- **Peer-assessment:** feedback on development and expression of ideas
- **Teacher-assessment:** effectiveness of letter, language use

e.g., Integrated performance assessment in US schools

Consortium for assessing performance standards
http://flenj.org/CAPS/?page=parent

Why these tasks?
1. Which tasks do we assess?

Which tasks?

→ Based on the U.S. Standards for Foreign Language Learning, by specific age/grade level and proficiency
→ 3 integrated tasks—Interpretive, interpersonal, presentational—reflect distinct modes of performance
→ Tasks are closely integrated with what is being taught: this is why you are learning what you are learning
→ Tasks provide the opportunity for self-, peer-, and teacher-assessment and feedback
→ Assessment raises awareness about effectiveness of teaching and learning

See Adair-Hauck, et al. (2006)
2. How do we elicit task-based performance?

*e.g. : Georgetown German: Prototypical Performance Writing Task*
Level IV course, “Text in Context”

The European Union and the United States:
Possible Comparisons and Lessons

Task
You have been invited to give a lecture at the Society for German-American Relations in Weimar next month. For your contribution, the organizers are looking for a presentation on the much-discussed topic “The European Union and the Founding of the United States - A Model?” Your talk is limited to 15 minutes, to be followed by discussion of your remarks.

Please observe your position as a speaker to this particular audience. While you should expect good background knowledge from your German audience regarding the situation in the EU, you primarily contribute the perspective of an American who is familiar with the foundational history of the USA and subsequent development of the country. In its particularities this is most likely to be less familiar to your audience. Therefore, it is precisely those aspects that you must bring to the awareness of your audience in order to link them to the present situation in Europe. In other words, continuously keep in mind this occasion for your speech, this starting point for your presentation, and this intention of your talk.
Example: CNaVT
Certificate in Dutch as a Foreign Language

Tourist domain
Listening comprehension task
“Visit to the museum”

Item: Identify the paintings based on the descriptions you hear from an audio museum guide.

What is the purpose of the paintings?
2. How do we elicit task-based performance?

e.g.: assessing young/beginning learners, US schools

Task Title: Recycling
Theme: Environment
Level: Novice-Mid (Focus Age Group: 11 - 14)
Communicative Mode: Presentational

Time Frame: One 42 minute period or more depending on the class and the type of presentations chosen.

Description of Task:
Now that you’ve learned about saving the planet and the need to reuse and recycle, you want to share with your family your ideas for reusing some throw away things. You have cleverly thought of a few uses for reusing ___________. Now you will share your new ideas. Create and develop your idea and choose one from the following types of presentations to present it.
2. How do we elicit task-based performance?

Title: ¿Qué puedo hacer con_________?
(What can I do with_______?)

Include the following in your presentation:
• four or more ways of re-using something
• specific uses
• descriptions/elaboration

Three options, you choose:
➔ Power point presentation
➔ Poster
➔ Live oral 3-D presentation

Materials Needed: computers, markers, poster board, realia

Why give the student these options?
2. How do we elicit task-based performance?

**Task-based performance elicitation?**

- Give a purpose for communicating (motivate)
- For interactive/presentational tasks, indicate or provide an audience
- Provide a structure or procedures for doing the task
- Utilize authentic input and other realia for situating the task in a meaningful context
- Indicate desired outcomes or products of the task
- Where feasible, encourage ownership over the task through individual choice in its accomplishment
3. What gets assessed in task performance?

**Test task:** Order pizza for the class.

Task accomplishment?

- Take orders
- Place orders
3. What gets assessed in task performance?

**Test task:** Order pizza for the class.

- Take orders
- Place orders

Task accomplishment?
3. What gets assessed in task performance?

e.g.: classroom assessments for middle-school learners in Flanders (Jordens, 2007)

**Type task: (from learning goals)**

Pupils can report to the teacher on a subject, previously discussed at school.

**Test task:**

Pupils report to the teacher on a film about scuba diving, the report being based on photographs extracted from the film.
3. What gets assessed in task performance?

Parameters for assessment

What does pupil say?
Content relevance, completeness
Clarity
Organization

How does pupil say it?
Fluency
Articulation
Manner of speaking
Variation and extensiveness
Accuracy: word/sentence
Interaction skills

### Parameters for assessment

<table>
<thead>
<tr>
<th>7. uitspraak</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Heel helder</td>
<td>o 0 binnensmonds praten</td>
</tr>
<tr>
<td>□ Helder</td>
<td>o 0 lispelen</td>
</tr>
<tr>
<td>□ Niet zo helder</td>
<td>o 0 letters inslikken</td>
</tr>
<tr>
<td>□ Niet helder</td>
<td>o 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8. spreukrant</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Heel aangenaam</td>
<td>o 0 te stil</td>
</tr>
<tr>
<td>□ Aangenaam</td>
<td>o 0 te luid</td>
</tr>
<tr>
<td>□ Niet zo aangenaam</td>
<td>o 0 te hoog</td>
</tr>
<tr>
<td>□ Niet aangenaam</td>
<td>o 0 te laag</td>
</tr>
<tr>
<td>□ 0 vlakke intonatie</td>
<td>o 0 overdreven intonatie</td>
</tr>
<tr>
<td>□ 0 hees/schor</td>
<td>o 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9. gevarieerdheid (relevant bij filmtaken?)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Heel gevarieerd</td>
<td>o 0 dezelfde woorden keren terug</td>
</tr>
<tr>
<td>□ gevarieerd</td>
<td>o 0 'dinges'</td>
</tr>
<tr>
<td>□ Niet zo gevarieerd</td>
<td>o 0 Steeds dezelfde zinsconstructie (bv en dan, en dan, en dan...)</td>
</tr>
<tr>
<td>□ Niet gevarieerd</td>
<td>o 0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10. correctheid</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WOORDNIVEAU</td>
<td>o 0 systematische fout op</td>
</tr>
</tbody>
</table>
3. What gets assessed in task performance?

Give a presentation in Spanish about your life and interests. Indicate how these activities reflect your personality. Include some possibilities for your future in terms of career, study, travel, pursuing personal interests.

Source:
http://depts.washington.edu/mellwa/Events/20081105/sandrock_ipa_handout.pdf
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Exceeds Expectations</th>
<th>Meets Expectations</th>
<th>Does Not Meet Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Do we understand you?</strong> (Comprehensibility)</td>
<td>My message is easily understood by my audience.</td>
<td>I am able to communicate my message by reproducing memorized material.</td>
<td>I rely heavily on visuals to convey my message.</td>
</tr>
<tr>
<td><strong>How do I use the Spanish language?</strong> (Language Control &amp; Vocabulary Use)</td>
<td>I can write/speak in simple sentences correctly. My presentation is rich in appropriate vocabulary.</td>
<td>I demonstrate some accuracy in written/oral presentation when I reproduce memorized words and phrases. My vocabulary reveals basic information.</td>
<td>My presentation is correct only at the word level. My vocabulary is limited and / or repetitive. I use English.</td>
</tr>
<tr>
<td><strong>How do I organize the presentation?</strong> (Communication strategies)</td>
<td>My presentation is supported with many enriched examples.</td>
<td>My presentation has some visuals/realia and examples.</td>
<td>My presentation lacks organization.</td>
</tr>
<tr>
<td><strong>How well do I impact the audience?</strong> (Impact)</td>
<td>I use visuals/ realia and gestures, and tone of voice to maintain audience’s attention.</td>
<td>I use some visuals/realia and gestures to maintain audience’s attention. My tone of voice is acceptable.</td>
<td>I use a small amount of effort to maintain audience’s attention.</td>
</tr>
</tbody>
</table>
Criteria for judging performance?

→ Consider real-world criteria for task success or accomplishment, including as appropriate:
→ Performance products: Does something happen, does something get produced?
→ Performance steps: Do critical elements of the task happen?
→ Performance qualities: linguistic and paralinguistic features (e.g., accuracy, fluency, gesture, pragmatics), content (e.g., meaning), other attributes (e.g., visual support, efficiency of task completion)
→ Performance levels: How well does the task performance meet expectations, standards, etc.?
4. How do we score task performance?

e.g., Georgetown Univ. German writing assessments

End-of-level Prototypical task

Formative assessment: Process writing

Summative assessment: Performance writing
4. How do we score task performance?

- **Formative assessment:** Process writing
  - Providing rich feedback to enable learning

- **Task-based assignment**
- **Level scoring rubric**

- **Students**
  - Writing task v. 1
  - Rating rubric, In-text
  - Initial Grade

- **Final Grade**
  - Writing task v. 2
  - Initial Grade
4. How do we score task performance?

Example Journalistic portrait: The Vietnamese immigrant experience in contemporary Germany

End-of-level Prototypical task

2 raters from the curricular level

+ 2 raters from another level

Providing reliable judgments of language + task abilities

Summative assessment: Performance writing
4. How do we score task performance?

Scoring and score reporting for different purposes?

→ Consider how ‘scores’ will be used before deciding how to score:
→ Formative assessment: Use public criteria and scales, and encourage self-, peer-, teacher-scoring for feedback
→ Summative assessment: Use well-defined criteria and scales, and multiple, trained raters in scoring for judgment
→ Holistic scoring (overall judgment) provides an indication of task success or accomplishment or completion
→ Analytic scoring (by categories and qualities) provides a useful basis for remediation
→ Rich and detailed score reporting enables learning from assessment (e.g., in person, score + justification)
Up Next:
The range of current uses for task-based language assessment

But First:
Additional Questions?
V. Workplace Tasks: Reading

--Scan basic charts, tables, maps or schedules for information. (5)
--Find routine information on the computer screen/scanner screen/computerized display screen, if available. (6)
--Read information in the reception/appointment book to find available openings for a new appointment. (6)
--Read a checklist to verify if all the steps in the procedure have been completed. (6)
--Follow one page of clear familiar instructions. (7)
--Read a reminder or complaint letter/memo, take appropriate action. (7)
--Scan complex charts, tables and schedules for several specific pieces of information for comparison/contrast. (7)
--Follow instructions on evacuation procedures, fire drills, or on using simple machinery/equipment. (7)
--Use plain language manual with familiar topic and content in own field of knowledge to find specific information. (8)

ETC...
V. Workplace Tasks: Reading

--Scan basic charts, tables, maps or schedules for information. (5)
--Find routine information on the computer screen/scanner screen/computerized display screen, if available. (6)
--Read information in the reception/appointment book to find available openings for a new appointment. (6)
--Read a checklist to verify if all the steps in the procedure have been completed. (6)
--Follow one page of clear familiar instructions. (7)
--Read a reminder or complaint letter/memo, take appropriate action. (7)
--Scan complex charts, tables and schedules for several specific pieces of information for comparison/contrast. (7)
--Follow instructions on evacuation procedures, fire drills, or on using simple machinery/equipment. (7)
--Use plain language manual with familiar topic and content in own field of knowledge to find specific information. (8)

ETC…

Who should be able to do these tasks?

Canadian Language Benchmarks:

English Language Demands for Nursing

---
Tasks as targets: Policies & standards

(e.g., Canadian Language Benchmarks)
“The CLB is task-based: • In syllabus design, tasks are considered to be basic building blocks. • Tasks in language learning promote the integration of all aspects of communicative competence, and multilevel language processing.”

Why?
→ Tasks provide meaningful targets: Ability to use the language
→ Washback on teaching and learning
4. **Integrated Speaking**: Read a passage from a psychology textbook and listen to the lecture that follows it. Then answer the question.

**Reading: Flow**

In psychology, the feeling of complete and energized focus in an activity is called flow. People who enter a state of flow lose their sense of time and have a feeling of great satisfaction. They become completely involved in an activity for its own sake rather than for what may result from the activity, such as money or prestige. Contrary to expectation, flow usually happens not during relaxing moments of leisure and entertainment, but when we are actively involved in a difficult leisure and enterprise, in a task that stretches our mental or physical abilities.
Listen to the Lecture (Male professor): I think this will help you get a picture of what your textbook is describing. I had a friend who taught in the physics department, Professor Jones, he retired last year. . . . Anyway, I remember . . . this was a few years ago . . . I remember passing by a classroom early one morning just as he was leaving, and he looked terrible: his clothes were all rumpled, and he looked like he hadn’t slept all night. And I asked if he was OK. I was surprised when he said that he never felt better, that he was totally happy. He had spent the entire night in the classroom working on a mathematics puzzle. He didn’t stop to eat dinner; he didn’t stop to sleep . . . or even rest. He was that involved in solving the puzzle. And it didn’t even have anything to do with his teaching or research; he had just come across this puzzle accidentally, I think in a mathematics journal, and it just really interested him, so he worked furiously all night and covered the blackboards in the classroom with equations and numbers and never realized that time was passing by.

Question: Explain *flow* and how the example used by the professor illustrates the concept.
Lecture (Male professor): I think this will help you get a picture of what your textbook is describing. I had a friend who taught in the physics department, Professor Jones, he retired last year. . . . Anyway, I remember . . . this was a few years ago . . . I remember passing by a classroom early one morning just as he was never felt better, that he was totally happy. He had spent the entire night in the classroom working on a mathematics puzzle. He didn’t stop to eat dinner; he didn’t stop to sleep . . . or even rest. He was that involved in solving the puzzle. And it didn’t even have anything to do with his teaching or research; he had just come across this puzzle accidentally, I think in a mathematics journal, and it just really interested him, so he worked furiously all night and covered the blackboards in the classroom with equations and numbers and never realized that time was passing by.

Question: Explain flow and how the example used by the professor illustrates the concept.
Tasks in large-scale proficiency assessment

(e.g., TOEFL iBT)
“Working directly with higher education institutions, we designed the TOEFL test to simulate actual academic classroom tasks.”

Why?
→ Tasks provide improved interpretations, construct validity
→ Washback on teaching and learning
Tasks in professional language certification

Example...
Tasks in professional language certification

Example...

Q: Who should be able to do these tasks?

http://www.anglo-continental.com/en/uk/forms/Aviation/Part%202%20ready.mp3
Tasks in professional language certification

For example...

International Civil Aviation Organization: English for Aviation Purposes

http://www.anglo-continental.com/en/uk/forms/Aviation/Part%202%20ready.mp3
Tasks in professional language certification

(e.g., ICAO, 2007)
“Because of the high stakes involved, pilots and controllers deserve to be tested in a context similar to that in which they work and test content should therefore be relevant to their roles in the workplace.”

Why?
- Tasks demonstrate ‘on-the-job’ abilities for high-stakes decisions
- Washback on teaching and learning

- English Aviation
- CLB-Nursing
- NCATE L2 teaching
- FBI Translation
- OET Medical
Overview of task

In groups, learners design a brochure for the Hong Kong Tourism Board describing 4 attractions in Hong Kong which would appeal to young people of their own age.

Writing a Tourist Brochure

Imagine the Hong Kong Tourism Board has asked your class to design a brochure that would be of interest to young visitors of your own age. In groups of 4, design the brochure describing FOUR sites suitable to young people of your own age coming to Hong Kong. Complete this task by following the steps below:

Step 1: Group Task.
Discuss in your groups which sites young people would want to visit in Hong Kong. Choose one site each to investigate. For homework find out as much as you can about the site, where the site is, when it is open, what one can see / do there, what the facilities are, how one gets there, etc. Bring this information to the next class.
Step 2: Group Task.
Exchange information with your group members. Tell them about the site you have found out about. Then decide how you are going to present the information in your brochure, what order you want to put your sites in, what illustrations you need, what title you want to give the brochure, etc.

Step 3: Individual Task.
Write a description about your chosen site (120 words). Remember to say why it is interesting. Proofread it carefully, then hand it to your teacher.

Step 4: Group Task.
In your groups edit your work based on your teacher’s comments. Then put together your brochure. Your brochures will be assessed on the following basis:
• Task fulfilment: would your selected sites appeal to young people?
• Accuracy of language and information provided: is the brochure written in good English? Is the information provided accurate?
• Attractiveness of final written submission: is your brochure really attractive? Can you make it more appealing?
Step 2: Group Task.
Exchange information with your group members. Tell them about the site you have found out about. Then decide how you are going to present the information in your brochure, what order you want to put your sites in, what illustrations you need, what title you want to give the brochure, etc.

Step 3: Individual Task.
Write a description about your chosen site (120 words). Remember to say why it is interesting. Proofread it carefully, then hand it to your teacher.

Step 4: Group Task.
In your groups edit your work based on your teacher’s comments. Then put together your brochure. Your brochures will be assessed on the following basis:
• Task fulfilment: would your selected sites appeal to young people?
• Accuracy of language and information provided: is the brochure written in good English? Is the information provided accurate?
• Attractiveness of final written submission: is your brochure really attractive? Can you make it more appealing?
Tasks in classrooms and programs

Task-based assessment for English language learning at secondary level: One of the key principles underpinning the school English language curriculum in Hong Kong is that of task-based language teaching. In this package we introduce the features of task-based assessment (TBA) within a school-based curriculum...

http://cd1.edb.hkedcity.net/cd/eng/TBA_Eng_Sec/index.html

Local practice

TBLT research

Teacher guides

ACTFL IPAs

Can-do

e.g., DIALANG

Why?

→ Tasks enable feedback on learning, teaching, curriculum
→ Washback on teaching and learning
Task-based language assessment in practice

**Classroom:** TBLAssessment For Learning

- **Profession:** Tasks for certification of important abilities
- **Large-scale:** Tasks for demonstrating proficiency in societal domains (e.g., education)
- **Program:** Tasks as summative indicators of teaching effectiveness
- **Standards:** Tasks as valued targets for learning
Some challenges and opportunities
In Task-Based Language Assessment...
Basic validity questions for TBLA

• How do we interpret task-based L2 performance?
  → authentic criteria, proficiency, L2 production?
• Can task-based performances be rated reliably?
  → analytic/holistic rating scales, rater training?
• What is the nature of task difficulty?
  → cognitive complexity, learner individual diffs?
• Can we generalize and/or extrapolate based on tasks?
  → domain-sampling, ECD, underlying constructs?
• How does task-based assessment impact instruction?
  → teacher, learner, curricular consequences?

“Fundamentally...assessments are only good insofar as their use does good, in terms of supporting educational efforts and outcomes, and it is my intent that validity evaluation helps to ensure that they do.”

Norris (2008)
How do we interpret task-based L2 performance?

- Can do a specific task
  - High-fidelity task simulation or replication
    - Test of aviation English

- Has a certain proficiency
  - Domain sampling, ‘type’ tasks
    - TOEFL, CNAVT

- Displays X language qualities
  - Implicit, spontaneous elicitation
    - Classroom practice, SLA research
Can task-based performances be rated reliably?

YES!!!

Individual task accomplishment: single task

- Highly specified indigenous criteria, steps, etc.
- Teacher, tester, learner awareness-raising

Ability for L2 use: multiple/type tasks

- Cross-task holistic/analytic rating scales
- Rater training
What is the nature of task difficulty?

Task features
e.g., number of elements, number of steps, planning time, structure, reasoning requirements, code frequency/complexity

Why is this task so hard?

Task conditions
e.g., number of participants, open/closed task, convergent/divergent, familiarity, gender, status

Individual differences
e.g., task familiarity, motivation, testing stakes, aptitude, personality

More difficult = better or worse performance???
Can we generalize and/or extrapolate based on tasks?

Of course we can → and we do it all the time!

Driving test → series of tasks (1-time!)

Generalization → can drive a car

Extrapolation → can negotiate a roundabout in London???

...be careful!
Can we generalize and/or extrapolate based on tasks?

1. Listening comprehension?
2. Listening comprehension in a particular domain?
3. Listening comprehension on a specific task?

1. Adequate construct definition
2. Adequate domain sampling
3. Adequate task representation
How does task-based assessment impact instruction?

Local impact (Georgetown German):

“So I think it has filtered down to smaller writing assignments. First of all so that students get used to it and it’s a coherent system. Whereas before we would say, ‘Okay, now write one page about your family and your Heimat’. Now we actually make up a sheet and use the same features.”

Global impact (TOEFL)

“...considerable changes occurred in the amount of attention the teachers paid to the development of speaking and to the integration of different skills.”
In sum, task-based performances and assessments may not (should not) meet all purposes for assessment within language education, but they probably offer a lot of advantages for certain purposes:

- **Task-based performances** tell us things that other item types do not (e.g., ability to accomplish particular tasks under particular conditions)
- **Task-based performances** offer rich opportunities for eliciting and observing language use with meaningful content in context (there is always more to language use than just language, and tasks capture that)
- **Task-based performances** can reveal multiple aspects of language ability and/or development within a single task (e.g., accuracy, complexity, fluency; content knowledge; procedural knowledge; expertise)
- **Task-based performances** are in many cases the most indicative targeted expectations of language classes and programs (if not, then what exactly are we teaching?)
- **Task-based assessments** provide relevant and comprehensive frames of reference for learners and teachers to generate feedback on form-function-meaning relationships (formative, learning-oriented uses)
- **Task-based assessments** communicate to teachers and learners and others what is valued within a given educational setting (motivational, educative uses)
- **Task-based assessments** may impact curriculum and educational systems, causing a rethinking of what and how teaching is occurring (washback uses)
- **Task-based assessments** generate interpretable data regarding the worthwhile outcomes of language learning programs (summative uses, e.g., for accountability, certification, program evaluation)
Why task-based assessment?

Q: Task-based assessment is a lot of work, and we can make somewhat accurate assessments of learners’ abilities and knowledge with short-cut tests; why should we bother with TBLA?

Tasks are useful:
- align to instructional practices and values
- provide rich samples of language use data
- inform how well we teach and learn
- show what L2 users can do, and motivate them to learn

Task-based assessments can:
- improve accuracy of interpretations
- ensure abilities to use language for specific purposes
- guide educational practices
- counter negative washback of other assessments
- enable a variety of uses, decisions, actions
Thank you!