

# Useful and practical assessment in the Arabic language classroom

Michael Raish, PhD

April 28, 2022

[LinkedIn](#)



*Mosque of Ibn Tulun. Photo credit: [Diego Delso](#), via Wikimedia Commons. [Image URL](#)*

## Plan for today

- Introductions
- Useful terms and concepts in assessment
- An example of an alternative Arabic test format
- Moving toward performance- or task-based assessment
- Recommendations for classroom practice
- Further reading and bibliography



*Hibiscus schizopetalus*. Photo credit: [Cayambe](#). [CC BY-SA 3.0](#), via Wikimedia Commons. [Image URL](#)



# Setting the (conceptual) stage

## Why do we use tests in our classrooms? [\(Douglas, 2010; p. 1-2\)](#)

Teachers spend a lot of time with their students—they can often provide accurate information about a student's current status and recent progress.

Tests provide:



Abuja National Mosque. Photo credit: [Mark Fischer](#), [CC BY-SA 2.0](#),  
via Wikimedia Commons. [Image URL](#).

## Why do we use tests in our classrooms?(Douglas, 2010; p. 1-2)

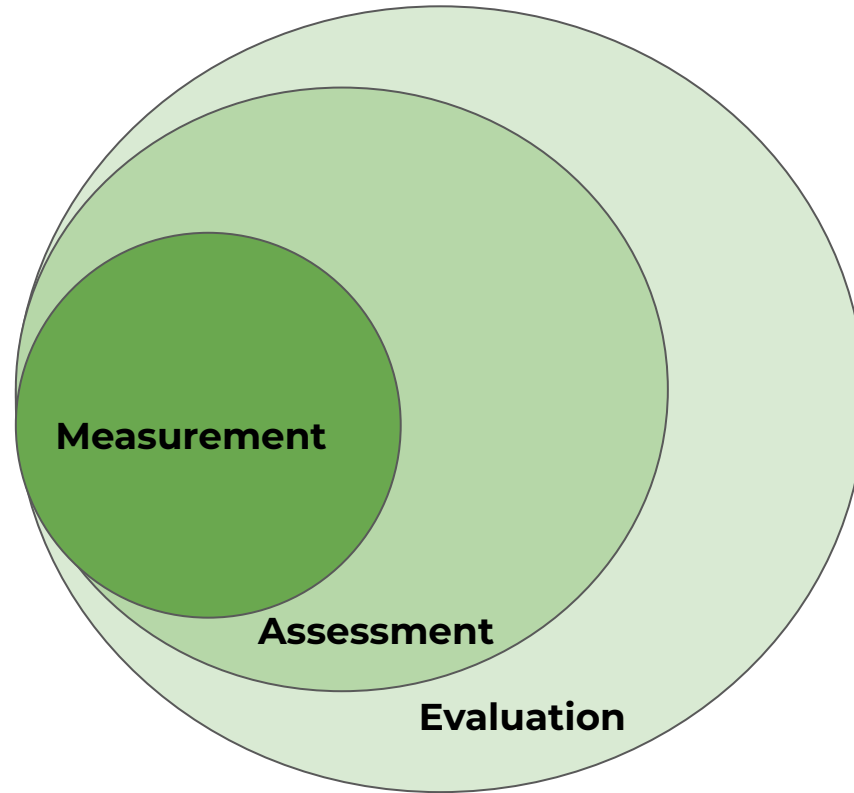
Teachers spend a lot of time with their students—they can often provide accurate information about a student's current status and recent progress.

Tests provide:

- Fairness and objectivity;
- Same instructions, same input, same conditions;
- “Second opinion” on teacher intuition;
- Standardization—compare students to each other and to past/future cohorts;
- Allow other **stakeholders** to access progress toward standards;
- **Achievement tests** demonstrate mastery of material;
- **Proficiency tests** demonstrate overall language ability;
- Tests are important to community/state/national policy.



Abuja National Mosque. Photo credit: Mark Fischer, [CC BY-SA 2.0](#),  
via Wikimedia Commons. [Image URL](#).



## Measurement is...

*the process of quantifying the characteristics of persons according to explicit procedures and rules.* (Bachman, 1990, p. 18)

### measurement relies on:

<i>Quantification</i>	Assigning numbers (not qualitative)
<i>Characteristics (traits; constructs)</i>	Physical or mental (language testing)  Quantification of <i>aptitude, intelligence, motivation, fluency, achievement</i> , etc.
<i>Rules and procedures</i>	Follows rules;  Orderly, not haphazard or random;

 Anyone can rate another's speaking, but under what conditions can this rating be considered "measurement"?

## Assessment is...

*a process of collecting information about something that we are interested in, according to procedures that are systematic and substantially grounded.* (Bachman, 2004, p. 7)

*the systematic gathering of information about student learning in support of teaching and learning.* (Norris, 2006, p. 579)

An important form of feedback for learners and teachers, [as advocated by ACTFL](#).

## Categorizing assessment<sup>(modified from Tran, 2012; p. 7)</sup>

<i>Mode</i>	oral or written
<i>Intention</i>	informal or formal
<i>Purpose</i>	formative or summative
<i>Interpretation</i>	Norm-referenced <ul style="list-style-type: none"><li>- Compare to other test-takers; e.g., SAT</li></ul> Criterion-referenced <ul style="list-style-type: none"><li>- Compare to body of knowledge; e.g., certification exam</li></ul>
<i>Administration</i>	classroom-based or state-/nation-wide

 **What do we want to know, and why do we want to know it? Who will use the information, and what actions will be taken on its basis?**



## Evaluation is...

*the process of using the measurements gathered in assessments . . . to judge the relationship between what was intended by the instruction and what was learned . . . to determine what students know and understand, how far they have progressed and how fast, and how their scores and progress compare to those of other students.* ([STU online](#))

Table  
adapted  
from  
**Richards**  
(2001).

formative	summative	illuminative
Provide students and teachers with information about the learning process.  Ongoing; formal or informal  <i>How are materials or objectives working?</i>	Determine the <b>effectiveness, efficiency,</b> and other aspects of the <b>performance of a language program.</b>  <i>What did students learn? Were the methods and materials effective?</i>	Gain a <b>deeper understanding of the learning/teaching process</b> without making major changes.  <i>How are students dealing with this kind of text? Is my feedback being received and taken up?</i>

## Types of test items

Selected response	Constructed response
Receptive	Productive or productive+receptive (e.g., after watching a video or reading a text)
Choose from multiple correct options	Students provide correct answer—speaking or writing.
Easy to score and administer—lots of students at once	Harder to score and administer
Wide range of knowledge	“Authentic” performance
Hard to design and validate	Hard to design and validate

### examples

- |   |  |
|---|--|
| <ul style="list-style-type: none"><li>• Multiple choice</li><li>• Matching</li><li>• True/false</li></ul> | <ul style="list-style-type: none"><li>• Fill-in-the-blank</li><li>• “Proficiency” tests</li><li>• Essays</li></ul> |
|---|--|



**What are some reasons that make selected response tests hard to design?**

## Proficiency is...

*... what individuals can do with language in terms of speaking, writing, listening, and reading in real-world situations in a spontaneous and non-rehearsed context.* (ACTFL, 2012, p. 3)

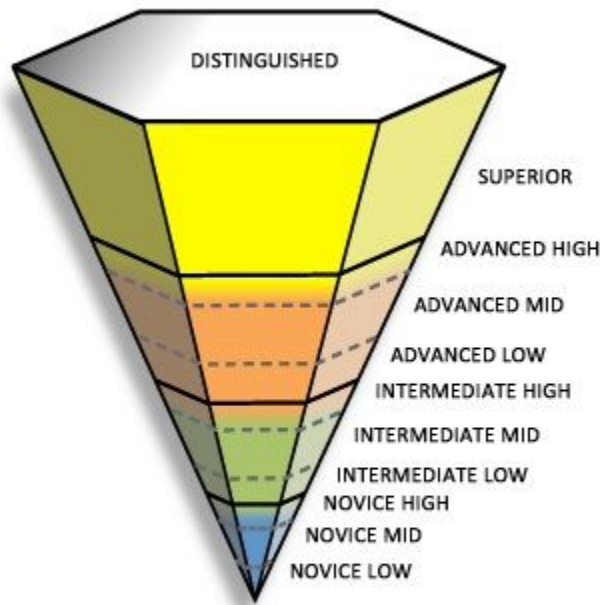
**The ACTFL proficiency guidelines and proprietary OPI have had a major positive impact on Arabic teaching in the United States.**

ACTFL OPI positive “washback effect” on the TAFL field:

- Communicative emphasis,
- “Can-do” statements,
- Performance-based,
- Assessment tasks resemble real-life and instructional tasks.

**However,**

- There are multiple ways to define “proficiency”.
- Guidelines lack empirical basis. (Liskin-Gasparro, 2003; Fulcher, 1995)
- Ratings are affected by test-taker anxiety. (Young, 1986; MacIntyre et al., 1997)
- Different raters may rate the same performance differently. (Brown, 2003)



ACTFL scale image reproduced from [General Preface to the ACTFL Proficiency Guidelines 2012](#). Accessed 4/26/22.

## Proficiency is also...

A combination of BLC and HLC. (Hulstijn, 2011; Hulstijn, 2018)

	Basic Language Cognition	Higher Language Cognition
<i>populations</i>	Native speakers are equal; learners differ.	Native speakers differ.
<i>domains</i>	Oral	Oral + Literacy
<i>awarenesses</i>	Implicit and unconscious	Explicit and conscious
<i>structures</i>	Morphology and syntax	Low-frequency vocabulary and uncommon grammar

 To what extent does the BLC-HLC distinction apply to native speakers of Arabic?

# Is there a “best” test of Arabic?

The [Center for Applied Linguistics](#) database lists 29 free and commercial tests of Arabic.

## Arabic [\[ edit \]](#)

- The Arabic Language Proficiency Test (ALPT)<sup>[3]</sup> is a standardized Arabic proficiency test designed by Arabic Academy and endorsed by the Islamic Chamber of Commerce and Industry (ICCI). There are 56 member countries under ICCI that cover all 22 Arab countries as well as 34 countries in Africa and the Far East. The ALPT started in 2002 and measures the student's level of proficiency in Arabic. The ALPT has five sections: Listening Comprehension, Reading, Structure, Writing, Speaking.
- Berlitz Language Institutes offer a competency test in reading, writing, speaking & listening assessing the proficiency in the Arabic language according to the CEFR benchmark.
- Eton Institute offers its own "Arabic Language Competency Test" (ALCT), a 4-skills (reading, writing, speaking, listening) exam which generates a band-score result similar to the IELTS model, but for Arabic.
- E-TOAFL and Al-Arabiyyah tests by the AL\_ARABIYYAH-INSTITUTE, are tests conducted online that measures a learner's level according to the CEFR benchmark.
- ACTFL [OPIc](#)
- CIMA (Certificat International de Maîtrise en Arabe)<sup>[4]</sup> is a standardized Arabic proficiency test developed by the Institut du Monde Arabe in Paris, and is offered since 2018 in Paris (at the Institute) and in several accredited centres around the world.
- Avant Arabic Proficiency Test (APT)<sup>[5]</sup> is a monolingual language proficiency test in Modern Standard Arabic (MSA) created for native Arabic Speakers who need to function effectively in the Arabic speaking world.

Screenshot of the [List of language proficiency tests](#) English Wikipedia article. Accessed 4/26/22.

John M. Norris (ed.)

Developing C-tests  
for estimating proficiency  
in foreign language research



PETER LANG

# An example of a “short-cut” Arabic proficiency test

## [Link to handout](#)

### **A. Introduce ourselves**

### **B. Questions for your group:**

1. Which of these texts is “harder”? How do you know?
2. What levels of instruction would the original texts be appropriate for?
3. What are they testing? What does someone’s score tell you?
4. What testing scenarios might this instrument be applicable to?

## عائلي

اسمي حسن، وأنا مصري من مدينة دمياط في شمال مصر. لي عا \_\_\_\_\_ كبيرة  
 فـ \_\_\_\_\_ الولايات المتـ \_\_\_\_\_ الأمريكية وفـ \_\_\_\_\_ مصر أيـ \_\_\_\_\_ .  
 لي خا \_\_\_\_\_ ، تسكن فـ \_\_\_\_\_ مدينة هيوستين حـ \_\_\_\_\_ يعمل زو \_\_\_\_\_  
 في شر \_\_\_\_\_ بترول، ولهـ \_\_\_\_\_ خمسة أو \_\_\_\_\_ ، ولدان و3 بنـ \_\_\_\_\_ .  
 ولدها الكبـ \_\_\_\_\_ يعمل ضا \_\_\_\_\_ في الجـ \_\_\_\_\_ الأمريكي والصغـ \_\_\_\_\_  
 متخصص فـ \_\_\_\_\_ الكيمياء. والبـ \_\_\_\_\_ الأولى تدـ \_\_\_\_\_ الأدب  
 العر \_\_\_\_\_ في جا \_\_\_\_\_ نيويورك والثا \_\_\_\_\_ اسمها ليلي ولـ \_\_\_\_\_  
 تعمل الأ \_\_\_\_\_ لأن عندها طفل جديد وتسكن في مدينة دالاس حيث يعمل زوجها  
 والثالثة متخصصة في علوم الكمبيوتر. لي أيضا أبناء خال يسكنون في ولاية  
 نيو جيرسي حيث يعملون في مطعم عربي.

## قوات الأمن المصرية تقتحم جامعة الأزهر بالقاهرة

اقتحمت قوات الأمن المصرية جامعة الأزهر في القاهرة حيث كان يتظاهر طلاب  
 مؤيدون للرئيس المعزول محمد مرسي. وأطلقت قو \_\_\_\_\_ الشرطة الغـ \_\_\_\_\_  
 المسيل للدمـ \_\_\_\_\_ من أ \_\_\_\_\_ تفريق المتظا \_\_\_\_\_ .  
 وقد قا \_\_\_\_\_ السلطات إ \_\_\_\_\_ الهدف مـ \_\_\_\_\_ اقتحام الجا \_\_\_\_\_ هو  
 موا \_\_\_\_\_ أعمال الشـ \_\_\_\_\_ . وأشارت إ \_\_\_\_\_ "تمكُن الأجـ \_\_\_\_\_ الأمنية  
 مـ \_\_\_\_\_ إحكام السيـ \_\_\_\_\_ على محـ \_\_\_\_\_ المبنى الإد \_\_\_\_\_ للجامعة  
 وتأ \_\_\_\_\_ رئيس الجا \_\_\_\_\_ ومسؤوليها وموظ \_\_\_\_\_ المحتجزين بدا \_\_\_\_\_ .  
 وكان طلـ \_\_\_\_\_ الجامعة يتظ \_\_\_\_\_ منذ أسا \_\_\_\_\_ احتجاجا عـ \_\_\_\_\_ عزل  
 الرئيس المنتخب محمد مرسي، في شهر يوليو/تموز الماضي على يد الجيش، عقب  
 احتجاجات شعبية واسعة. ويطالب المتظاهرون بعودة ما يصفونه بالشرعية.

1. Which of these texts is "harder"? How do you know?
2. What levels of instruction would the original texts be appropriate for?
3. What are they testing? What does someone's score tell you?
4. What testing scenarios might this instrument be applicable to?



# قوات الأمن المصرية تقتحم جامعة الأزهر بالقاهرة

اقتحمت قوات الأمن المصرية جامعة الأزهر في القاهرة حيث كان يتظاهر طلاب مؤيدون للرئيس المعزول محمد

مرسي. وأطلقت قو \_\_\_\_\_ الشر \_\_\_\_\_ من أ \_\_\_\_\_ تفريق المتظاهرين.

- lexis
- morphology (plural)
- morphology (feminine)
- morphology (non-human)

وقد قا \_\_\_\_\_ السلطات إ \_\_\_\_\_ الهدف م \_\_\_\_\_ اقتحام الجا \_\_\_\_\_ هو موا \_\_\_\_\_ أعمال الش \_\_\_\_\_.

وأشارت إ \_\_\_\_\_ "تمكّن الأج \_\_\_\_\_ الأمنية م \_\_\_\_\_ إحكام السي \_\_\_\_\_ على مح \_\_\_\_\_ المبنى

الإد \_\_\_\_\_ للجامعة وتأ \_\_\_\_\_ رئيس الجا \_\_\_\_\_ ومسؤوليها وموظ \_\_\_\_\_ المحتجزين بدا \_\_\_\_\_." وكان

طل \_\_\_\_\_ الجامعة يتظ \_\_\_\_\_ منذ أسا \_\_\_\_\_ احتجاج \_\_\_\_\_ ب محمد مرسي، في

- lexis
- morphology (plural)
- morphology (feminine)
- morphology (human)
- morphology (possession)
- syntax (case) + morphology

شهر يوليو/تموز الماضي على يد الجيش، عقب احتجاجات \_\_\_\_\_ بعودة ما يصفونه

بالشرعية.

# قوات الأمن المصرية تقتحم جامعة الأزهر بالقاهرة

30 أكتوبر/ تشرين الأول 2013

اقتحمت قوات الأمن المصرية جامعة الأزهر في القاهرة حيث كان يتظاهر طلاب مؤيدون للرئيس المعزول محمد مرسي.

وأطلقت القوات الغاز المسيل للدموع لتفريق المتظاهرين.

وقالت السلطات إن الهدف من اقتحام الجامعة هو مواجهة أعمال الشغب.

وأشارت إلى "تمكن الأجهزة الأمنية من إحكام السيطرة على محيط المبنى الإداري لجامعة الأزهر وتأمين رئيس الجامعة ومسؤوليها وموظفيها المحتجزين بداخله."

ودعت الحكومة إلى "عودة الأمن" للجامعات حتى لو اقتضى الأمر إعادة الحرس الجامعي الذي ألغى بحكم قضائي.

 What level of Arabic instruction is this text appropriate for?

# Features of a C-test

## Context

- First and last sentences left intact, providing important context

## Lots of items

- 25 deletions per text
- 5 texts of varying difficulty

## “Random” deletions naturally target range of structures

- Delete half the letters of every other word

## Arabic works best with a word-final deletion strategy

- Definite article ال not counted
- Word-final morpho-syntactic structures are often deleted
  - Feminine marker ة
  - Interaction of syntax (case) with morphology (humanness); e.g., موظفها

## advantages

- easy to administer (lots of tests at same time)
- measure multiple kinds of knowledge/structures
- multiple texts of varying difficulty
- highly adaptable to different curricula and testing contexts

## disadvantages

- time-consuming to develop and sequence
- literacy-bound
- accesses varied *capacities*, but not necessarily *proficiency*—i.e., a “short-cut” proxy for overall Arabic ability



# Toward performance- or task-based assessment

# Lineage of Arabic instruction in the United States

## Grammar-translation

- Philological tradition of Arabic as a “Classical” or liturgical language<sup>([Ryding, 2017](#))</sup>

## Army method

- Sudden need of capable interpreters and translators—emphasis on reading and oral production<sup>([Richards & Rodgers, 2014, p. 58](#))</sup>
- Led to adoption of *audio-lingual method* in academic institutions.
- Arabic first taught by American Army Language School after WWII.<sup>([Nielsen, 2011](#))</sup>

## Arabic Proficiency Test <sup>([Rammuny 1975; 1983; 1999](#))</sup>

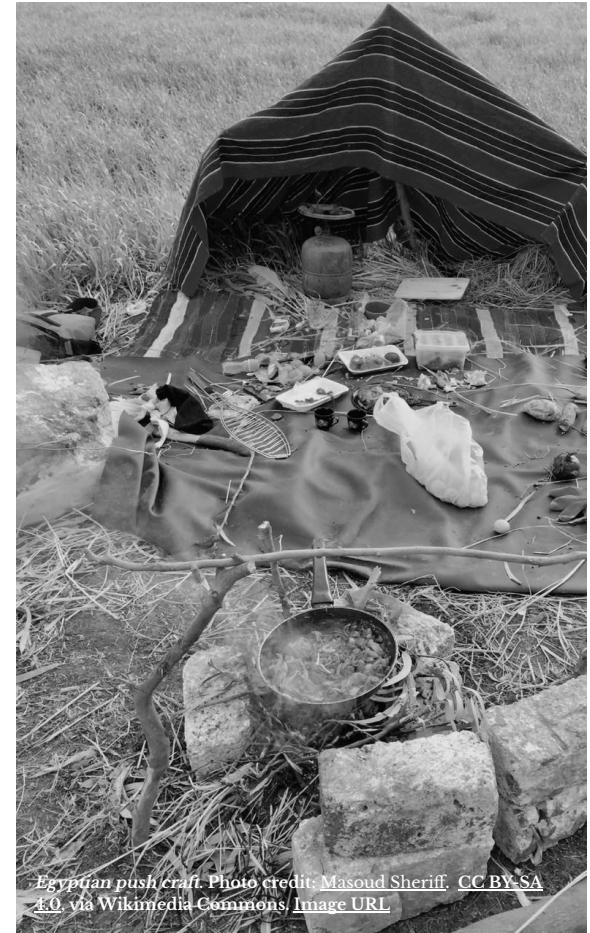
- Declarative grammatical and lexical knowledge
- Years of study is not a reliable indicator of “proficiency”

## Communicative Competence<sup>([Canale & Swain, 1980](#))</sup>

- Grammatical knowledge is only a *part* of competence. To what extent can learners actually *use* the language in a contextually appropriate manner?

## ACTFL OPI

- Lack of tools to assess large numbers of trained language professionals.
- Emphasis on “proficiency”-oriented instruction and assessment for Arabic.<sup>([Barhoum, 1989](#))</sup>



*Egyptian push/craft.* Photo credit: [Masoud Sherif](#). CC BY-SA 4.0 via [Wikimedia Commons](#)/Image URL

# Task-Based Language Teaching

TBLT organizes curriculum, instruction, and assessment around the concept of “task”; i.e., **real-life or real-world linguistic activities and situations** that (a) bear relationship to the world outside the classroom, and that (b) have been identified as an outcome of instruction.

Task-based assessment accommodates learning objectives and directly demonstrates learning and competencies related to *what a learner can do with the language*.<sup>(Skehan, 1998)</sup>

## Target tasks, aka what happens “out there”

- What should students be able to do after instruction?
- Included in learning objectives.
- Likely to be encountered in students’ personal, professional, and/or academic lives.
- Focus of assessment—how well can students complete the target task?

## Pedagogic tasks, aka what happens “in here”

- **Based on the target task**—what are the linguistic components of successful target task completion?
- Activities and materials that form the basis of language instruction.
- Have a communicative goal.
- Can be successfully “completed” through communication.

# What is a task?

## Example target tasks

- Apply for a job
- Take notes on a lecture
- Order a meal at a restaurant
- Give directions over the phone to a delivery driver
- Write a school report
- Apply for a scholarship
- Decide among friends where to eat
- Give a presentation
- Give feedback to a peer on their presentation

## Example pedagogic tasks

- Discussions
- Problem solving
- Listening activities
- Decision making
- Peer review/peer feedback
- Pair discussion
- Interviews
- Essay writing
- Exchange of information necessary to complete the task



**Pedagogic tasks should emphasize learner production, be “authentic,” and be scored according to a rubric.**



**Production-oriented tasks allow us to elicit “*natural-like*” language. However, they allow *limited observations* and can be *time consuming* to design, administer, and score.**

## Why bother focusing on task-contextualized performance?

Because task-based assessments: (modified from Norris, 2016, pg. 241)

- Align well with an emphasis on functional language use;
- Tell us things that other types of tests can't—e.g., the ability to use different types of language under different conditions;
- Offer rich opportunities for eliciting, observing, and providing feedback on language use;
- Serve as useful representation of learning outcomes.

## Challenges to task-based assessment of Arabic

Performance or task-based assessment likely cannot (and should not) fulfil all assessment needs within a language program.

What is “authentic” task performance?

- Which linguistic forms are required?
- *How much* colloquial vs MSA is appropriate or desirable?
- What level of accuracy is acceptable?

How can “authentic” performance be replicated in the classroom?

Arabic often lack support and resources.



# Using tasks in assessment

## Curriculum designers and instructors should consider:

- What order should tasks come in? In other words, “task sequencing.”
- A. What is a typical developmental trajectory for Arabic learners at your level of instruction?
  - *Some broad trends have been established, but TAFL would benefit from more research.*
- B. Cognitive processing demands—which tasks are more complex, and which are less?
  - *Modify task complexity by repeating tasks, +/- planning time, +/- scaffolding, etc.*
- C. Which tasks are more or less difficult?
- D. What are the lexical, grammatical, and discourse expectations of this genre of production?
- E. What is the role of the instructor?
  - *Oral instructions*
  - *Directing journaling and journal sharing*
  - *Incorporating note taking*
  - *Demonstrating, or leading students through completion of the task in Arabic.*



**How much of the above is necessary to address for your day-to-day Arabic teaching?**

## Using tasks in assessment

Performance-based assessment offers an **alternative to traditional discrete-point tests** and shed more light on “real-world” language use.

Assessment is **responsive** to frequent changes in curriculum and instruction.

Instructors observe students perform the task. How well can they do it?

ACTFL advocates a [“backwards design” to planning assessment](#)—what are you going to test at the end of the unit or course?

**However, task-based assessment of Arabic requires a nuanced consideration of “accuracy”.**

### Example target tasks

- ***Ordering tea at a cafe***
- ***Writing an academic essay***
- ***Applying for a job via email***
- ***Giving directions to a taxi driver***



**Do any of the tasks at right present different requirements for grammatical accuracy? Lexical complexity?**

## From *curriculum* to *instruction* to *assessment*

Are we emphasizing our students' *use of Arabic* to accomplish authentic tasks?

• • •

If so, discrete-point tests of *declarative Arabic knowledge* may be confusing or misaligned.

### An example of integrated performance assessment<sup>(adapted from Norris, 2016, p. 238)</sup>

- A. Students read about employment in the Arab world *in Arabic*, followed by a *self assessment* of their own knowledge and ability to linguistically complete the task.
  - *Self-assessments can raise awarenesses of learners and teachers, and highlight gaps in competencies.*
- B. Students discuss what they've learned *in Arabic* with a partner or in a small group, and provide systematic, rubric-aligned *feedback* on their partners' performance.
  - *External or peer feedback provides an important perspective on the learning process.*
- C. Students complete a job application and/or application letter for an Arabic-using workplace.
  - *The highly structured written documents allow teachers opportunities for formalized, systematic feedback.*

**Some  
things to  
think about  
for the  
classroom**

*Mustansiriyah Madrasah.* Photo credit: [David Stanley](#). CC BY 2.0, via Wikimedia Commons. [Image URL](#).

# Things to think about when designing pedagogic (classroom) tasks<sup>(adapted from Fulcher & Davidson, 2007; pp. 69-70)</sup>

## Interactional activity: Who has the info?

- Participants hold different information and need to exchange it to complete the task;
- One participant holds all information;
- Each participant holds all information.

## Interaction requirement: Who needs the info?

- Each participant requests and supplies information;
- One participant supplies all information;
- Each participant is expected to request and supply information but is not required to do so.

## Communication goal: Who wants to do what?

- Participants have the same goals;
- Participants have separate goals.

## Outcome options: When is the task “complete”?

- Only one outcome is “success”
- Multiple outcomes are possible



- Partners or groups—different roles
- (1) gives directions to a destination, (2) attempts to follow on own map.
- Success = arrival at destination
- +/- planning time
- +/- shared objectives
- +/- information access

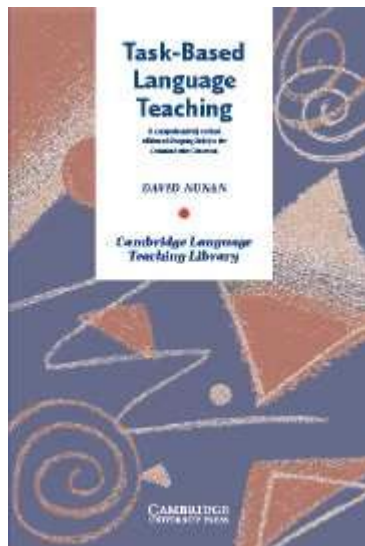
# General best practices in classroom test design

Identify not only *what* you want to measure, but *how* you plan to use that information.

1. Identify *knowledge, skills, or abilities* that you want to test
  - *What situations are these behaviors used in? How are they used? What linguistic structures form part of successful, authentic performance?*
2. Connect tests and test items to course learning outcomes and objectives.
3. Connect tests and test items to classroom instruction—*test what you teach*; i.e., don't diverge too far from classroom practice.
4. Be clear and explicit with students—e.g., have them participate in task and test design; help them understand how and why the test is being used.
5. To the extent possible, use authentic, “natural” language.



## Further reading and bibliography



Nunan, D. (2010). *Task-based language teaching*. Cambridge University Press.

- American Council on the Teaching of Foreign Languages (ACTFL). (2012). *ACTFL Proficiency Guidelines 2012*. ACTFL. <https://www.actfl.org/sites/default/files/guidelines/ACTFLProficiencyGuidelines2012.pdf>
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. OUP Oxford.
- Bachman, L. F. (2004). *Statistical Analyses for Language Assessment*. Cambridge University Press.
- Barhoum, K. (1989). Proficiency Testing: Rating vs. Profiling. *Al-Arabiyya*, 22(1/2), 63–80.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1–25.
- Brown, J. D., & Hudson, T. (1998). The Alternatives in Language Assessment: Advantages and disadvantages. *University of Hawai'i Working Papers in ESL*, 16(2), 79–103.
- Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Multilingual Matters.
- Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1), 1–47.
- Carroll, J. B. (1968). The Psychology of Language Testing. In A. Davies (Ed.), *Language Testing Symposium: A Psycholinguistic Approach* (pp. 46–69). Oxford University Press.
- Douglas, D. (2010). *Understanding language testing*. Hodder Education.
- Fulcher, G. (1996). Invalidating validity claims for the ACTFL oral rating scale. *System*, 24(2), 163–172. [https://doi.org/10.1016/0346-251X\(96\)00001-2](https://doi.org/10.1016/0346-251X(96)00001-2)
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge.
- Hulstijn, J. H. (2011). Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment. *Language Assessment Quarterly*, 8(3), 229–249. <https://doi.org/10.1080/15434303.2011.565844>
- Hulstijn, J. H. (2018). An Individual-Differences Framework for Comparing Nonnative With Native Speakers: Perspectives From BLC Theory: (Near)nativeness. *Language Learning*, 69, 157–183. <https://doi.org/10.1111/lang.12317>
- Liskin-Gasparro, J. E. (2003). The ACTFL Proficiency Guidelines and the Oral Proficiency Interview: A Brief History and Analysis of Their Survival. *Foreign Language Annals*, 36(4), 483–490. <https://doi.org/10.1111/j.1944-9720.2003.tb02137.x>
- MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in Self-Ratings of Second Language Proficiency: The Role of Language Anxiety. *Language Learning*, 47(2), 265–287. <https://doi.org/10.1111/0023-8333.81997008>
- Malone, M. E. (2011). *Assessment Literacy for Language Educators* (p. 2). Center for Applied Linguistics. [https://ez.cal.org/cal\\_user/content/download/1516/15923/file/AssessmentLiteracyforLanguageEducators.pdf](https://ez.cal.org/cal_user/content/download/1516/15923/file/AssessmentLiteracyforLanguageEducators.pdf)
- Nielsen, H. L. (2011). Second Language Teaching. In L. Edzard & R. de Jong (Eds.), *Encyclopedia of Arabic Language and Linguistics*. Brill Online.
- Norris, J. (2006). Development and evaluation of a curriculum-based German C-test for placement purposes. In R. Grotjahn (Ed.), *Der C-test: Theorie, empirie, anwendungen/the C-test: Theory, empirical research, applications* (pp. 45–83). Peter Lang.
- Norris, J. M. (2016). Current Uses for Task-Based Language Assessment. *Annual Review of Applied Linguistics*, 36, 230–244. <https://doi.org/10.1017/S0267190516000027>
- Nunan, D. (2010). *Task-based language teaching*. Cambridge University Press. <http://0-www.ebooks.cambridge.org/catalogue.uoc.edu/ebook.jsf?bid=CBO9780511667336>
- Ramunty, R. M. (1975). Arabic Proficiency Test. *Al-Arabiyya*, 8(1/2), 93–97.
- Ramunty, R. M. (1983). Arabic Proficiency Test: Implementation and implications. *Al-Arabiyya*, 16(1/2), 85–96.
- Ramunty, R. M. (1999). Arabic language testing: The state of the art. *Al-Arabiyya*, 32, 157–189.
- Richards, J. C. (2001). *Curriculum development in language teaching* (Nachdr.). Cambridge Univ. Press.
- Richards, J. C., & Rodgers, T. S. (2014). *Approaches and Methods in Language Teaching*. Cambridge University Press.
- Ryding, K. (2017). Teaching Arabic in the United States II. In *Handbook for Arabic Language Teaching Professionals in the 21st Century, Volume II: Vol. II* (pp. 11–19). Routledge.
- Schmitt, N. (Ed.). (2010). *An introduction to applied linguistics* (2nd ed). Hodder Education.
- Skehan, P. (1998). *A Cognitive Approach to Language Learning*. OUP Oxford.
- St. Thomas University Online. (n.d.). *Educational Measurement, Assessment and Evaluation*. St. Thomas University Online. Retrieved April 25, 2022, from <https://online.stu.edu/articles/education/educational-measurement-assessment-evaluation.aspx>
- Tran, T. H. (2012). *Second Language Assessment for Classroom Teachers*. 1–23.
- Young, D. J. (1986). The Relationship Between Anxiety and Foreign Language Oral Proficiency Ratings. *Foreign Language Annals*, 19(5), 439–445. <https://doi.org/10.1111/j.1944-9720.1986.tb01032.x>

وشكراً

**Thank you!**

**Questions?**

**[mraish@wikimedia.org](mailto:mraish@wikimedia.org)**



## Essays

- A reliable way of eliciting productive language use.
- Revealing of the learning process—lexical complexity appears to increase in tandem with “proficiency” among learners.

### Arabic-related challenges to essay tasks

- *Validity*
  - What does this written performance reveal about the learning process?
  - What are we measuring? How are we scoring, and for what purpose?
  - Are these written tasks “authentic” representations of Arabic use outside the classroom?
- *Modality*
  - Should we be handwriting or typing? How would a NS perform this task?
- *Practicality*
  - Time consuming to administer and reliably score.

## Essays should be:

- Focused—i.e., not a “free write.”
- Clear about the desired structure and length.
- Used more for formative than summative or achievement purposes.
- Consist of several short items, rather than a few long ones, if used for summative/achievement purposes.
- Have clear rubrics for measurement (to mitigate unconscious bias).

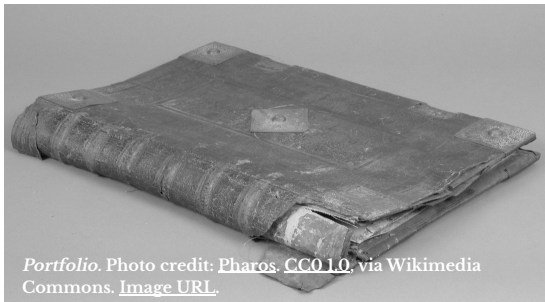
# Student portfolios—demonstrating progress over time or a course of study

## Benefits

- Ongoing/long-term.
- Flexible modality—physical, digital, or both.
- Facilitate collaboration between student and teacher.
- Highlight students' creative work:
- Can reinforce learning, improve motivation, and highlight gaps.
- Allows the re-negotiation of the teacher's role—providing clarity and feedback in a new context.

## What to include?

- Tests and quizzes,
- Homework,
- Projects,
- Recordings,
- Diary entries,
- Self assessments,
- Comments from peers.



Portfolio. Photo credit: [Pharos](#), CC0 1.0, via Wikimedia Commons. [Image URL](#).

## Challenges

- How should they be scored or graded?
- What types of artifacts should be included, in what format, and at what point in the curriculum?
- How can the information gathered via portfolio assessment be triangulated with learning outcomes?