

*Developing LSP tests for
different contexts:
Exploring the usability of the
ALTE LSP Guidelines*

Dr. Koen Van Gorp, Michigan State University
Dr. Dina Vîlcu, Babeş-Bolyai University

Introducing ourselves

- Dr Dina Vîlcu



UNIVERSITATEA
BABEȘ-BOLYAI

- Dr Koen Van Gorp



taal:
unie

CNaVT

Certificaat Nederlands als Vreemde Taal



Overview

- Introduce yourself (Task 1)
- Two central issues in LSP
 - Specificity (Task 2, 3)
 - Rating performances (Task 4, 5 and 6)
- Use the input and Guidelines to help your think about designing your own LSP test
 - Purpose of the test and relevant task types (Task 3)
 - Rating criteria (Task 6)

TASK 1: Introduce yourself

- Introduce yourself
- What is your interest in LSP?
- What (burning) question(s) do you have relating to LSP tests?

- Are you currently using an LSP test? Which one? (Standardized or created by you/team?)
- Are you planning to design an LSP test? For whom? What context?

Developing LSP: Guidelines

Guidelines for the Development of Language for Specific Purposes Tests

A supplement to the Manual for Language Test Development and Examining

Produced by ALTE



To help the language testing community with the development of LSP tests, the ALTE LSP SIG developed a practical, step-by-step guideline – as a supplement to:



Aim:

- More coherence
- Point of reference

Manual for Language Test Development and Examining

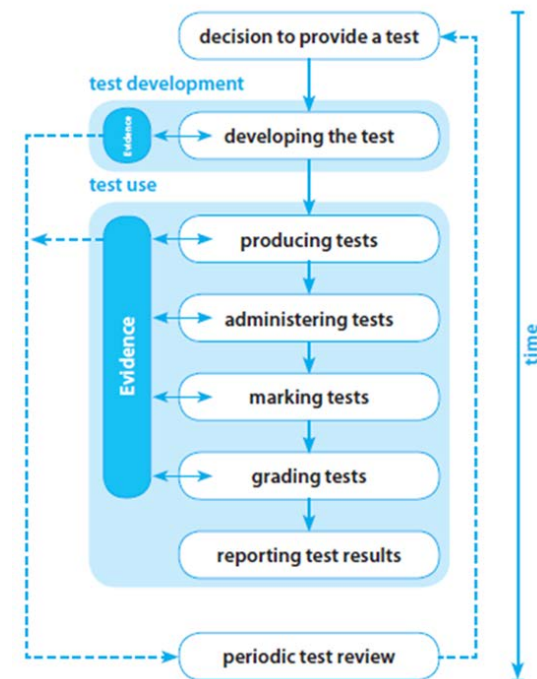
For use with the CEFR

Produced by ALTE on behalf of the Language Policy Division, Council of Europe



Structure of the Guidelines

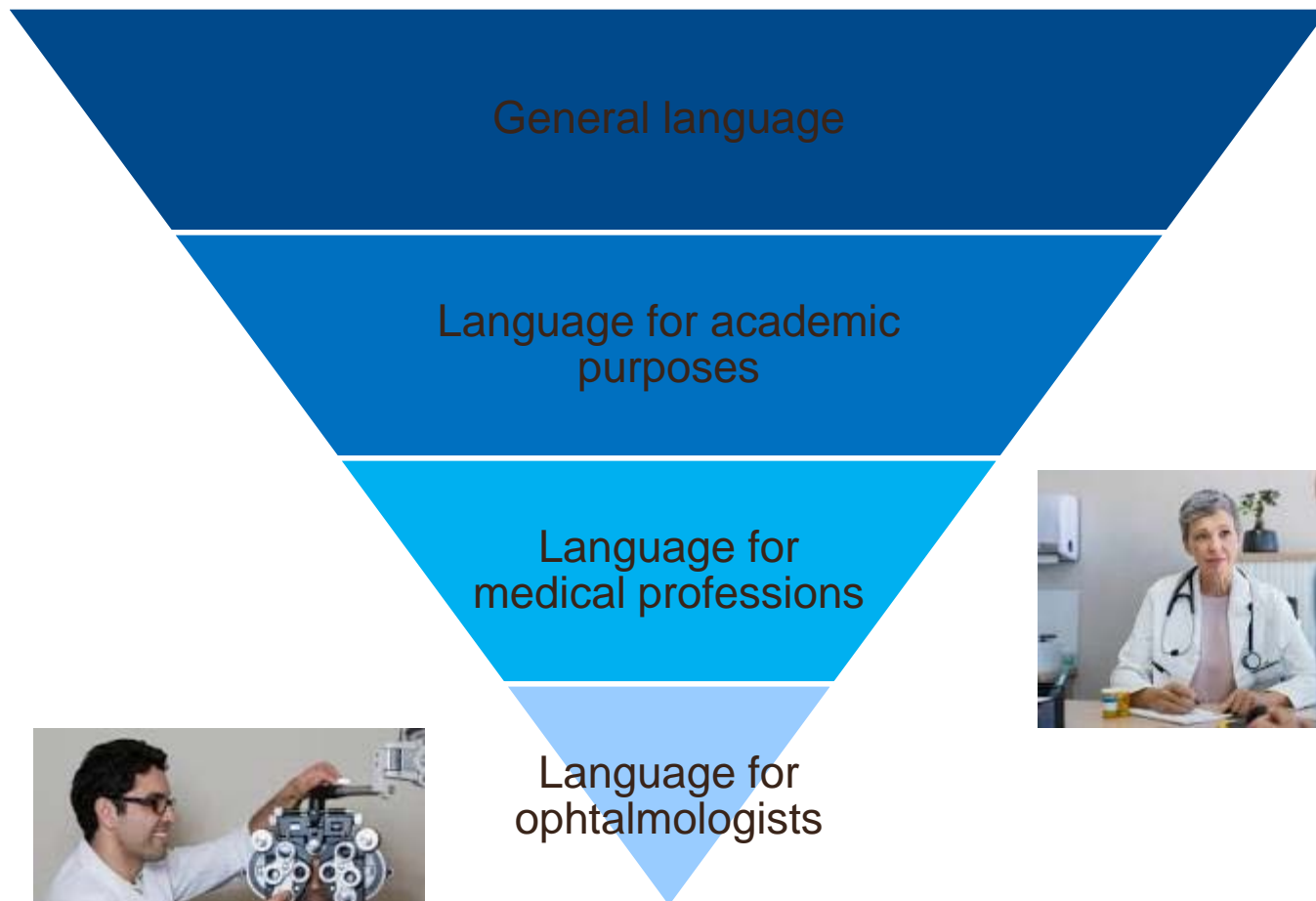
1. Developing the LSP test
2. Assembling LSP tests
3. Delivering tests
4. Marking, grading and reporting of results
5. Monitoring and review.



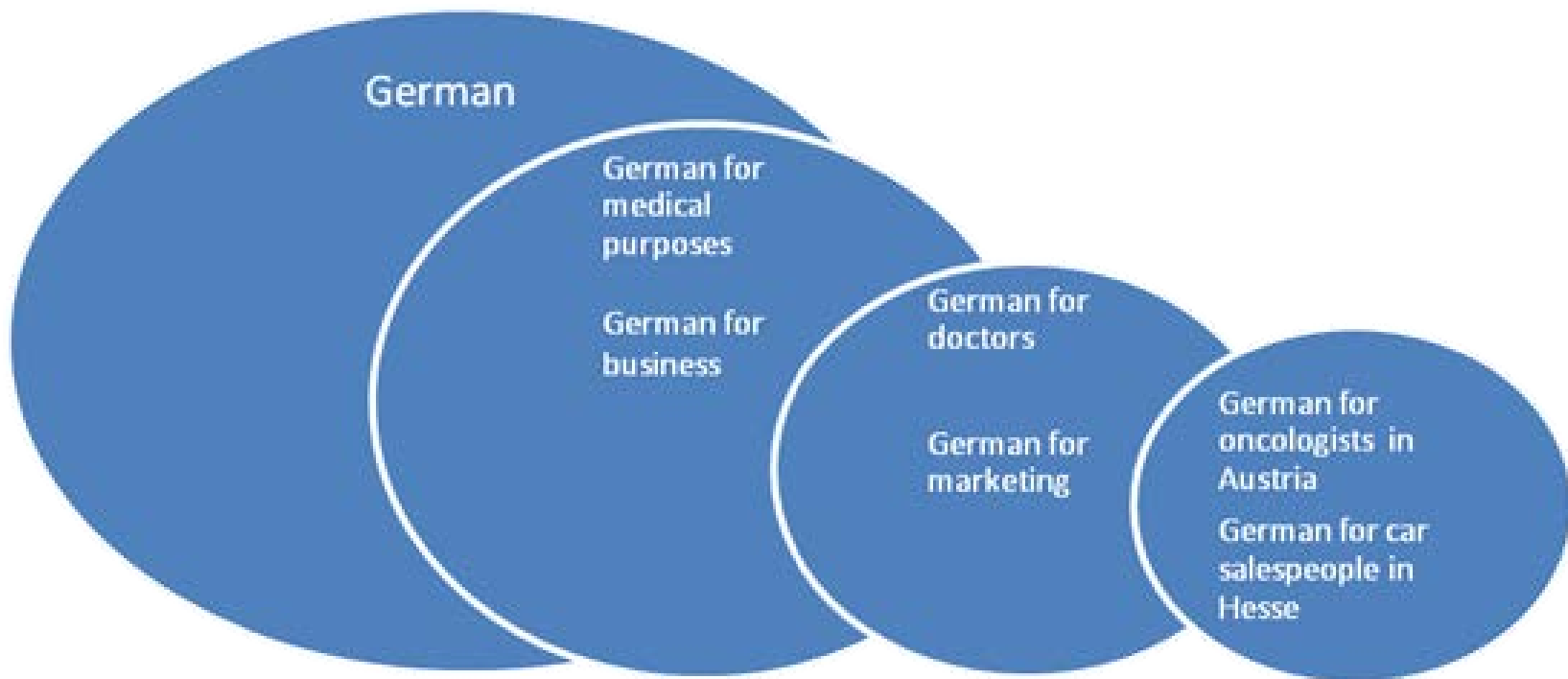
Issues in LSP

SPECIFICITY

Level of specificity

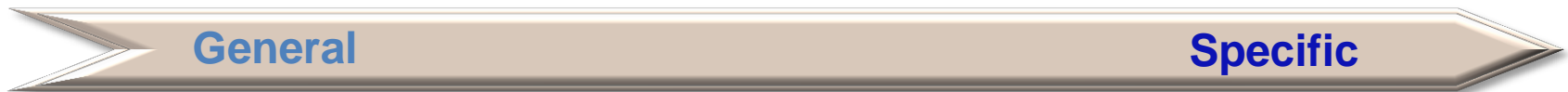


Level of specificity



How specific is specific?

Too much abstraction leads to test tasks that are too decontextualized to elicit the language behaviour appropriate to the domain...



it is clearly impractical to design a test for each situation or indeed, taking specificity to its logical conclusion, for each individual.

Aviation English

Routine landing approach to an airfield

Air Traffic Control Officer:

- *Fastair 345, cleared straight in ILS approach runway 28, descend to altitude 3000 feet QNH 1011, report established on the localiser.*

Pilot:

- *Cleared straight in ILS approach runway 28, descend to altitude 3000 feet, QNH 1011, Wilco, Fastair 345.*

Aviation English

Non-routine situations

(1) Pilot:

- *Control. Redline 253. We have a passenger on board with suspected heart attack. We'd like to divert to the nearest available airfield. Request full medical assistance on arrival.*

(2) Air Traffic Control Officer:

- *Genesis 1415. Understand you are having problems with your ailerons. Are you able to make left turns?*

(Emery, 2014: 200)

TASK 2

Health-related tasks from five tests:

- Which is which?
- How can you tell? How easy is it to identify them?
- What language abilities are required for each?

Source: Green, A. (2015). *Course on Assessing Language for Academic Purposes*. ALTE Workshop Bergen May 4-5, 2015.

5 health care tests

1. A test of **English for General Academic purposes**: the English language proficiency of people who want to study or work where English is used as the language of communication
2. A test for **medical professionals** intending to practice in the UK: the ability to apply knowledge to the care of patients
3. A **subject-specific admissions test** for applicants to medicine, veterinary medicine and similar courses at universities
4. A test of **English for medical professionals**: the language communication skills of healthcare professionals who wish to register and practise in an English-speaking environment
5. A test for **practising doctors for professional recognition**: advanced discipline- specific knowledge

Example 1

Findings from studies with animals suggest that caloric restriction could delay ageing in humans. Unfortunately, for maximum benefit, people would probably have to reduce their caloric intake by roughly 30 per cent, equivalent to dropping from 2,500 calories a day to 1,750. Few mortals could stick to that harsh a regimen, especially for years on end. But what if someone could create a pill that mimicked the physiological effects of eating less without actually forcing people to eat less?

YES *if the statement reflects the claims of the writer*

NO *if the statement contradicts the claims of the writer*

NOT GIVEN *if it is impossible to say what the writer thinks about this*

There is scientific evidence that eating less may extend human life. Not many people are likely to find a calorie-restricted diet attractive.

Example 2

Concerning the risk factors of uveitis associated with Juvenile Idiopathic Arthritis (JIA, Juvenile Rheumatoid Arthritis, JRA, or Juvenile Chronic Arthritis, JCA), which of the following factors increase the risk of developing uveitis?

1. The presence of Anti-Nuclear Antibodies (ANA positive).
2. An age of onset of arthritis between 5 and 10 years of age.
3. The presence of flare on slit lamp biomicroscopy.
4. Less than 5 joints are affected at the onset of the arthritis.

Example 3

The main argument of food manufacturers and government officials for the importance of the use of preservatives is that without them foods would soon spoil. This argument is indeed quite reasonable. However, it is interesting to note that of the nearly 4,000 different additives currently in use, over 90% are used purely for cosmetic reasons and as colouring agents, with preservatives accounting for less than 2% of all additives, by number or by weight.

Which one of the following statements is FALSE ?

1. Only a small percentage of additives are preservatives. B Ninety percent of food additives are colouring agents.
2. There are nearly 4,000 additives currently used in food.
3. The majority of food additives are non-essential.

Example 4

A 16 year old girl has had an enlarging mass in the right side of her neck for the last six weeks. She has had no other symptoms. She has a 2 x 2 cm enlarged lymph node in the anterior triangle of the neck with several smaller associated lymph nodes palpable. Oropharyngeal examination shows tonsillar membranes.

What is the SINGLE most likely diagnosis?

1. Infectious mononucleosis
2. Leukaemia
3. Lymphoma
4. Sarcoidosis
5. Tuberculosis

Example 5

Several large scale trials have shown, at a very high level of statistical significance, that those who take no alcoholic drinks are more likely to suffer from heart disease than those who drink moderate amounts of alcohol. The researchers have concluded that a moderate intake of alcoholic drinks protects the body against heart disease.

Which one of the following is a reason why this conclusion might be unsafe?

1. If enough trials are carried out, some will show statistical significance purely by chance.
2. Those people drinking no alcohol may have other factors in their diet or lifestyle which increases their susceptibility to heart disease.
3. A significant number of people refuse to answer questions about alcohol consumption.
4. By-products of alcohol production other than alcohol itself may be the reason for the protection against heart disease.

Key

Example	Test
1	1) English for General Academic purposes
2	5) Practising doctors for professional recognition
3	4) English for medical professionals
4	2) Medical professionals
5	3) Subject-specific admissions test

Identifying language needs

- What learners need to do with language
 - Can do statements
 - Specify the communicative tasks the test takers should be equipped to face.
- Tasks – Canadian Language Benchmarks (CLB)
 - A language task is a communicative “real world” instance of language use to accomplish a specific purpose in a particular context
 - Through these tasks learners demonstrate application of language knowledge (i.e., competence) and skill (i.e., ability)

Identifying tasks: B1 level

1. Answer the phone at work. Provide information to the caller (such as company name, hours of operation, basic information about a product or service).
2. Initiate a short, informal conversation with co-worker on how to use a system at work, such as the booking a travel or requesting a holiday, or how to transfer a call on a telephone system.
3. Give advice to a colleague about taking a day off because he/she is ill.
4. In a small group discussion, plan an event (such as a company open house, or benefit event). Discuss different options and come to an agreement.

Identifying tasks: B2 level

1. Answer the phone in a professional manner and respond to non-routine requests for detailed information about products or services.
2. Give instructions to a co-worker on how to process sales, handle merchandise returns, or operate a cash register.
3. Respond to a customer complaint, providing various suggestions to resolve it.
4. Participate in a performance review with an employer or instructor. Provide detailed information about successes achieved during the year.

TASK 3

- How specific do you want your (or your group's) LSP test to be? (See p.7-8 and 10 for Key questions)
- 1. Define purpose and test-taker population
- 2. Describe the specific domain
- 3. Identify some typical tasks that need to be performed in that domain
 - Handout: Canadian Language Benchmarks

Issues in LSP

RATING PERFORMANCES

TASK 4

- How would you rate the following three tasks?
 1. Writing task: Essay about Jung and Freud
 2. Speaking task: Buying eggs for a restaurant
 3. OET Speaking task

How could this task be scored?

Academic writing task

- Write an essay comparing and contrasting the views of Freud and Jung on the function of dreams

(Green, 2014: 136-137)

How could this task be scored?

Speaking task

- You are a purchasing agent for a wholesale catering business which sells eggs to restaurants. You need at least 240 large eggs every month, but you could buy up to 50 more as your business is expanding. The current price of a dozen eggs is typically 97 cents, but your director has told you that you must find a lower price. Delivery should be made within two days.
- Call three suppliers of eggs. Negotiate a discount.
- Place your order with the cheapest supplier who can meet your needs. (Green, 2014: 136-137)

OET Speaking task

- Specific to profession, based on typical workplace situations.
- One-to-one conversations with an interlocutor.
- A warm-up interview: candidate's professional background.
- Two role plays.
 - 2–3 minutes to prepare for each role play.
 - 5 minutes Role plays
 - Typical interactions between a health professional and a patient. The candidate adopts their usual professional role (e.g. as a nurse) and the interviewer plays a patient or, sometimes, a relative.
- <https://www.youtube.com/watch?v=j3llxOyoNk8>

(Rating of) Performance Testing

McNamara (1996) – continuum of performance testing

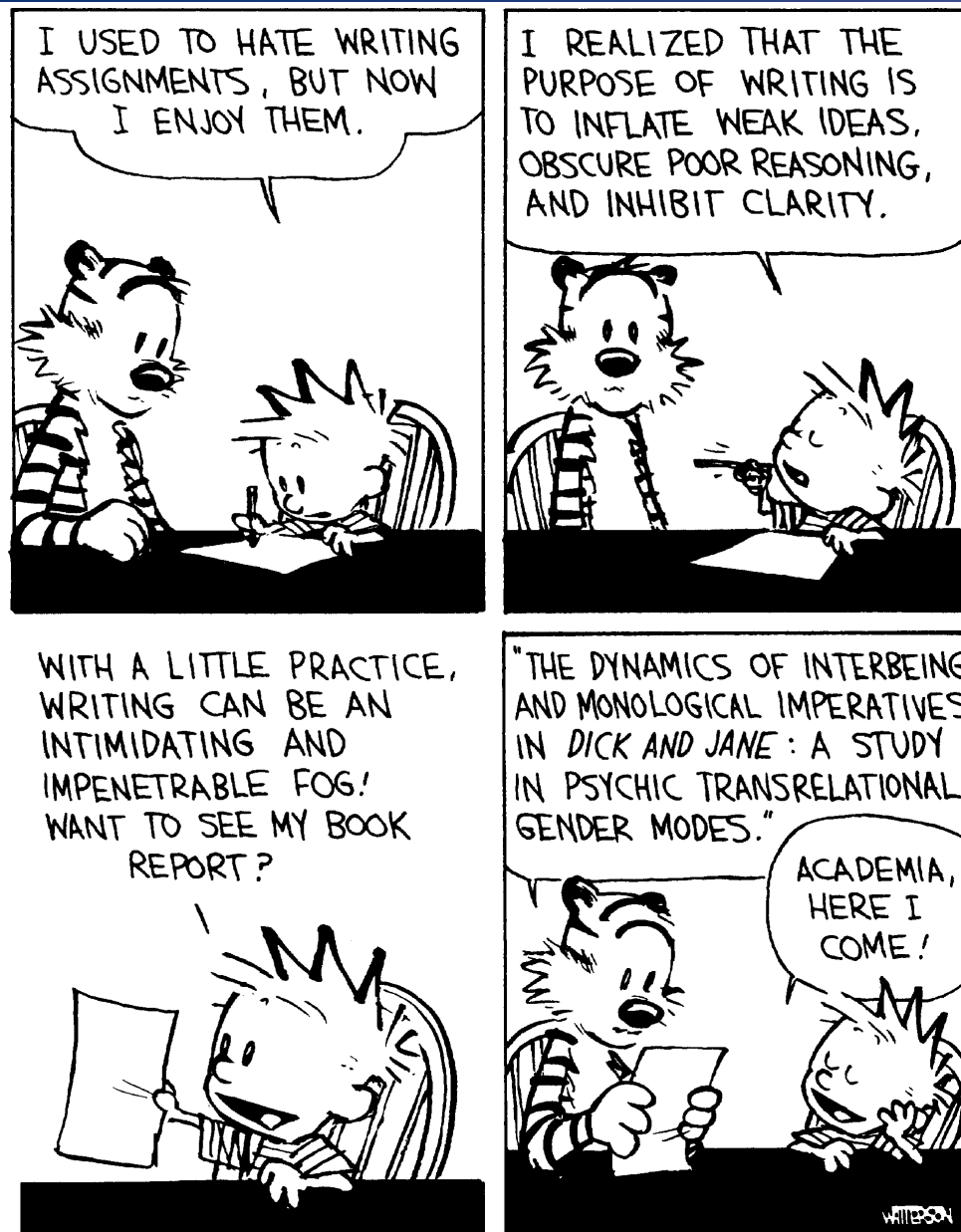
STRONG →

- *focus on successful task fulfilment*

←WEAK

- *focus on language used*

Looking at academic writing



Criteria for judging 'academic' writing

- Content understanding
- Criticality of judgement

- Rhetorical organisation
- Linguistic adequacy



***Subject
knowledge***

***Language
ability***

OET Speaking Assessment Criteria

Your performance on each of the two role-plays is scored against five criteria and receives a band score for each criterion:

1. Overall Communicative Effectiveness
2. Intelligibility
3. Fluency
4. Appropriateness
5. Resources of Grammar and Expression

However,

- Since the test [OET – role play] includes only a measure of the language used with no assessment of the quality of the medical advice given, the interactions are seen by test takers as being artificial and inauthentic. (O’Sullivan, 2012: 83)
- Alternatively: The Professional and Linguistic Assessment Board Test (for medical professionals) in the UK had language and medical professionals assess the same clinical assessment performances, each looking at a different aspect of the performance. (O’Sullivan, 2012: 83)



- “it is not uncommon for scales used in LSP assessment to establish a split between a range of generic language criteria such as lexical and grammatical knowledge or organization, and a task achievement criterion that is set out to represent the domain aspect.” (Brunfaut, 2014: 218)

‘Indigenous’ Assessment Criteria

- A plea for “indigenous” assessment criteria that are derived from a target language use analysis
- Criteria derived from subject specialists as they assess communicative performance (e.g. assessing the writing of a Physics report) are considered to be critically important in enhancing the construct validity of the LSP assessment.
(Douglas, 2000, 2001; Elder et al., 2012; Jacoby & McNamara, 1999)

LSP at Border Patrol Academy

Role-play tasks

- Critical scenarios with professional role-players
- Students role-play key course tasks over two days
- Criterion-referenced scoring
- Each scenario scored by two instructors



Burwell, Gonzales-Lloret & Nielson, 2009
Gonzales-Lloret & Nielson, 2015

LSP at BPA - Assessment

A. CRITICAL ACTIONS (3):

By the end of this module, students will complete the following critical actions. If the student successfully meets the success criteria during the scenario, place a check in the column marked "Yes." If the student does not demonstrate the success criteria (either through failure to perform or by not attempting the action), place a check in the column marked "No."

Action	Yes	No
A.1 Demonstrates Spanish language accuracy <i>Success criteria: Agent's commands, questions, and comments are appropriate to the situation.</i>		
A.2 Demonstrates Spanish language fluency <i>Success criteria: Pauses and/or misstarts in Agent's speech do not detract from comprehensibility.</i>		
A.3 Pronounces Spanish language clearly <i>Success criteria: Suspect is able to understand Agent's pronunciation and responds to commands and questions appropriately.</i>		

Sample
performance
-based
assessment
rubric

LSP at BPA - Assessment

B. KEY ACTIONS (3):

By the end of this module, students will complete the following key actions. If the student successfully meets the success criteria during the scenario, place a check in the column marked "Yes." If the student does not demonstrate the success criteria (either through failure to perform or by not attempting the action), place a check in the column marked "No."

Action	Yes	No
B.1 Inquires about subject's well-being <i>Success criterion: Agent asks subject about his/her well-being.</i>		
B.2 Ascertains subject's physical condition <i>Success criterion: Subject informs Agent of any health problems after questioning.</i>		
B.3 Determines need/desire for medical assistance <i>Success criteria: Agent arranges for emergency medical treatment or determines that emergency medical treatment is unnecessary.</i>		

Sample content-based assessment rubric

Developing assessment criteria

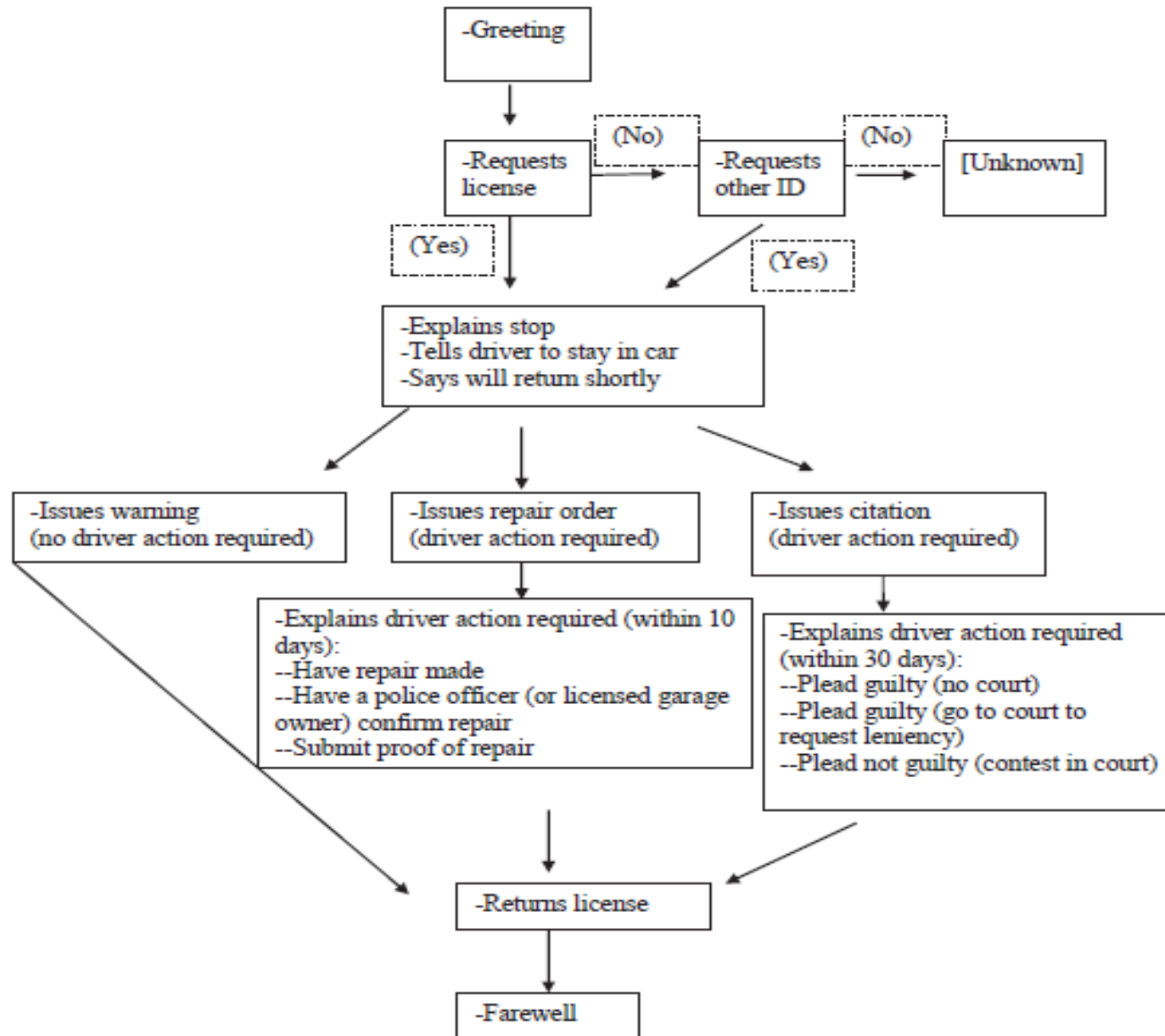
Consider **real-world criteria** for task success or accomplishment, including as appropriate:

1. *Performance products*: Does something happen, does something get produced? (task fulfilment)
2. *Performance steps*: Do critical elements of the task happen? (process evaluation)
3. *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)
4. *Performance levels*: How well does the task performance meet expectations, standards, etc.?

Police intervention: Traffic stop



Traffic stop intervention



(O'Connell, 2014: 124)

TASK 5

- Look at the two prototypical Traffic stop interventions (warning and citation) and at the schematic structure of the prototypical interaction.
- What assessment criteria would you use to rate these interactions?

TASK 6

- Consider the rating criteria we talked about.
- Create (potential) rating criteria for the LSP tasks that you identified relevant for your test.
 - See Chapter 5, p. 17-19

Some final questions, principles, thoughts

WRAPPING-UP

Key questions for LSP tests

- *Construct Definition*: How comprehensively is the language of the domain in focus defined or described? How is this done?
- *Test Relevance and Coverage*: How well do the test tasks reflect the target domain in terms of relevance and coverage?
- *Criterion Level*: If we accept that specific purpose tests should be criterion referenced, then how is the criterion arrived at? How is this level reflected in the test?
- *Scoring Systems*: Are the scoring systems appropriate to the domain?

Assessment and Instruction: Some principles

- Test what it is reasonable for the students to have been able to learn.
- Do not include test items which might require knowledge in the domain or cultural which have not been taught.
- Create integrated tasks, which involve both reception and production of text.
- Treat receptive and productive abilities equally.
- The test should contain exercises for checking domain specific vocabulary.

Specialists

- Build (language) assessment specialist 's understanding of the subject – create a shared frame
- Build subject specialist understanding of language assessment / linguistic terminology

Bridging both worlds: Context specialist, like teachers teaching LSP courses

Thank you!

Questions?