

Paper presented at ISB12

Symposium: Implication and Challenges of a National Framework for Bilingual Assessment in Canada

The CLB as a resource for test development and validation

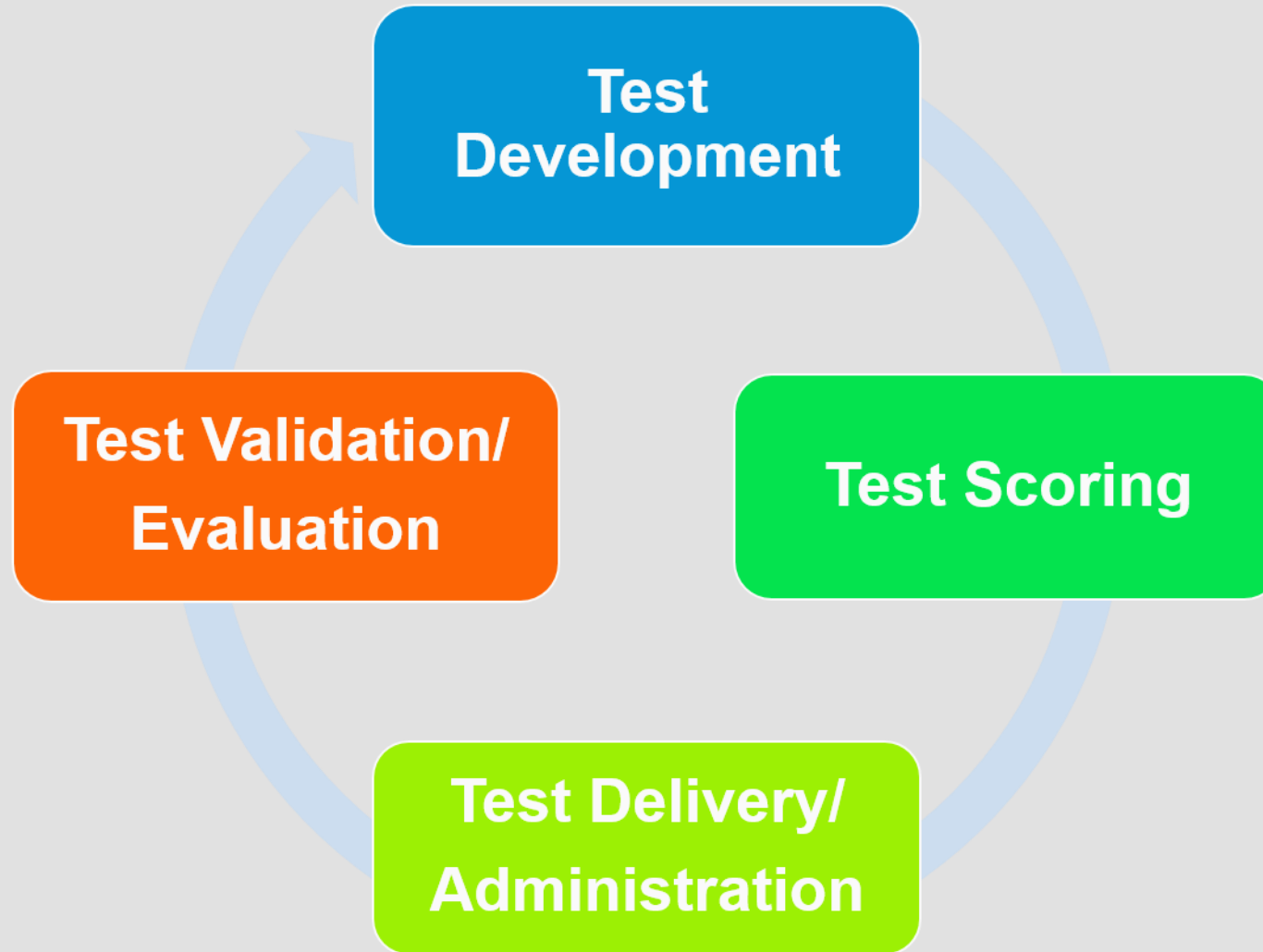
Michelle Y. Chen, Jennifer Flasko, & You-Min Lin

Paragon Testing Enterprises, Inc.
Vancouver, Canada



by  Paragon

Test Development, Operation, and Evaluation Cycle

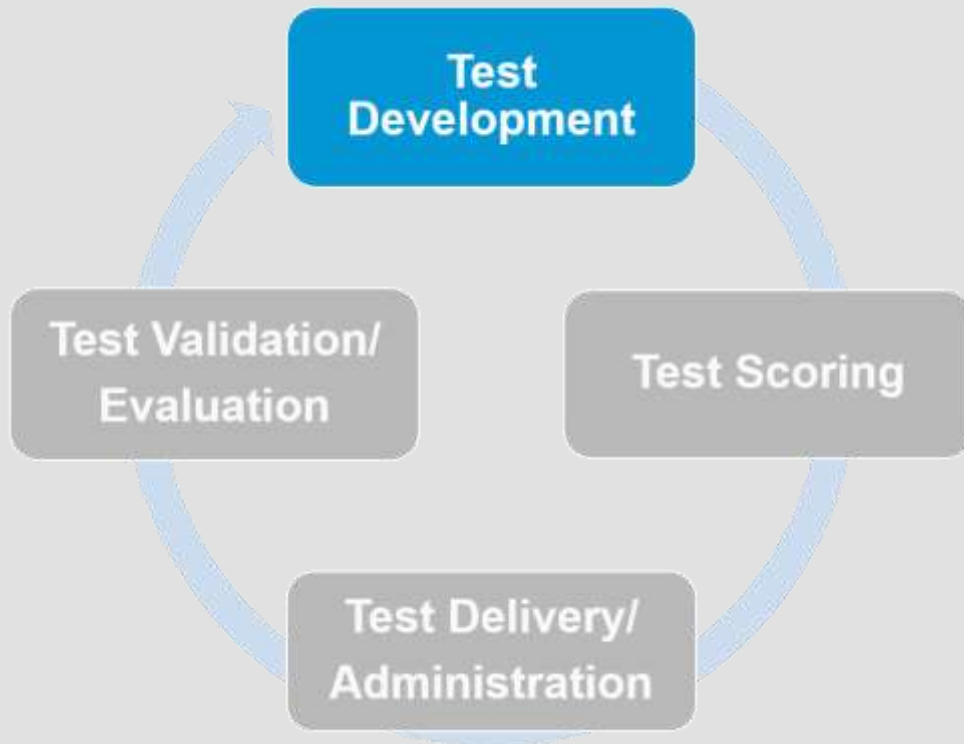


CELP-IP-General Tests



- Canadian English Language Proficiency Index Program General (CELPIP-General)
- Test of functional English proficiency for adults
- Primarily used for Canadian immigration purposes
- All computer delivered

Test Development Process



Domain Analysis	<ul style="list-style-type: none">• TLU Domain
Defining the Construct	<ul style="list-style-type: none">• Target Contexts• Language Knowledge & Skills
Designing the Specifications	<ul style="list-style-type: none">• Task Design
Pilot Testing	<ul style="list-style-type: none">• Test Prototype• Finalization

Test Development Process

Domain Analysis	<ul style="list-style-type: none">• TLU Domain
Defining the Construct	<ul style="list-style-type: none">• Target Contexts• Language Knowledge & Skills
Designing the Specifications	<ul style="list-style-type: none">• Task Design
Pilot Testing	<ul style="list-style-type: none">• Test Prototype• Finalization

Structure of this presentation:

Questions

What questions do we ask?

Activities

What activities do we undertake?

Outcomes

What do we aim to achieve?

Test Development Process

Domain Analysis	<ul style="list-style-type: none">• The TLU Domain
Defining the Construct	<ul style="list-style-type: none">• Target Contexts• Language Knowledge & Skills
Designing the Specifications	<ul style="list-style-type: none">• Task Design
Pilot Testing	<ul style="list-style-type: none">• Test Prototype• Finalization

Questions

1. Who are the target test takers?
 - What is the profile of the expected test taker population?
2. What are the features of the Target Language Use (TLU) Domain?
 - What kind of language is used, in what contexts, and by whom?

Test Development Process

Domain Analysis	<ul style="list-style-type: none">• The TLU Domain
Defining the Construct	<ul style="list-style-type: none">• Target Contexts• Language Knowledge & Skills
Designing the Specifications	<ul style="list-style-type: none">• Task Design
Pilot Testing	<ul style="list-style-type: none">• Test Prototype• Finalization

Activities

Research

- Demographic information (target test takers)
- Literature related to TLU domain (English in work, study, social contexts)
- CELPIP: newcomers to Canada
 - Participation in social, work & study contexts

Resources

- Immigration, Refugees & Citizenship Canada (IRCC)
- External proficiency frameworks & standards (CLB)

The Canadian Language Benchmarks (CLB)

- A national language standard for English
- A descriptive scale of English as a Second Language ability across a continuum
 - 3 stages: basic, intermediate & advanced
 - 12 reference points (“benchmarks”)
- Describes the progression of language knowledge & skills underlying basic, intermediate and advanced ability in adult ESL users



Canadian Language Benchmarks, October 2012 Edition, Introduction, pg. V

The Canadian Language Benchmarks (CLB)

Profiles of Ability (12 BMs across 3 Stages)	Knowledge & Strategies (Stages I - III)	Canadian Language Benchmark Pages (BM 1 – 12)
<ul style="list-style-type: none">• Statements of ability• Conditions of the context• Strengths & limitations	<p>Enabling background knowledge & skills for each stage:</p> <ul style="list-style-type: none">• Grammatical knowledge• Textual knowledge• Functional knowledge• Sociolinguistic knowledge• Strategic competence	<p>Profile of Ability</p> <p>Competency Areas (4)</p> <ul style="list-style-type: none">• Interacting with Others• Comprehending Instructions (R&L)• Giving Instructions (S)• Getting Things Done• Comprehending Information (R&)• Sharing Information (S&W)• Reproducing Information (W)

The Canadian Language Benchmarks (CLB)

The CLB is:

- “A framework of reference for learning, teaching, programming & assessing adult ESL in Canada”

The CLB is not:

- a curriculum no specific curriculum or syllabus
 - an assessment no test specifications
- ❖ “It is a standard that can inform assessment by providing the information needed to guide the development of assessment tools.”

Introduction, pg. V & XII

Canadian Language Benchmarks, October 2012 Edition,

The CLB as a Resource for Test Development



Domain Analysis

The CLB is a resource for...

- Helping test developers describe the domain of English use in work, study, and social contexts in Canada

Domain Analysis

- The TLU Domain

Defining
Constru

Outcome

1. Profile of the Target TT



Designi
Specific

2. Contexts of language use for newcomers in Canada

In their communities:

- interacting with retailers, service providers, government personnel, community members, etc.

At their workplaces:

- communicating with colleagues, attending meetings, participating in projects, etc.

Pilot Tes

Domain Analysis	<ul style="list-style-type: none"> • The TLU domain
Defining the Construct	<ul style="list-style-type: none"> • Target Contexts • Language Knowledge & Skills

Design
Specifi

Questions

Pilot Te

1. What areas of the TLU domain will be represented on the test?
 - Which contexts are most critical to test users?
2. What aspects of language will be assessed by the test?
 - What language knowledge & skills are most critical?

Domain Analysis	<ul style="list-style-type: none"> • The TLU Domain
Defining the Construct	<ul style="list-style-type: none"> • Target Contexts • Language Knowledge & Skills

Design Specific
Pilot Test

Activities

Construct Definition

- A formal statement of what the test intends to measure

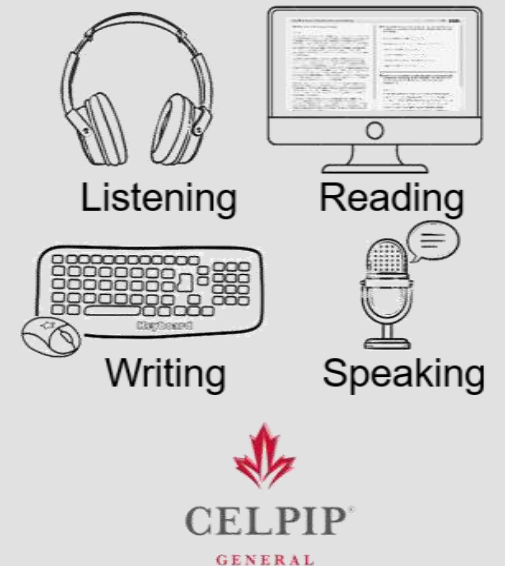
Resources

- Domain Analysis research

Outcome

Construct Definition (CELPIP)

“The CELPIP-General is a test of general English language proficiency. The test measures the functional Listening, Reading, Writing, and Speaking skills required for everyday communication. The test allows the test takers to demonstrate their English abilities in a variety of everyday situations, such as communicating with co-workers, interacting with friends, understanding newscasts, and interpreting and responding to written materials...”



CELPIP Theoretical Framework 2019 Edition, Paragon Testing Enterprises, Ltd.

Questions

1. What tasks best represent the TLU domain & test construct?

- critical contexts, language knowledge & skills

2. What is the desired length & delivery mode of the test?

Domain Analysis

Defining the
Construct

**Designing the
Specifications**

- **Task Design**

Pilot Testing

- Test Prototype
- Finalization

Activities

Designing the Test Tasks

Domain Analysis	<ul style="list-style-type: none">• Characteristics of the rubric• Characteristics of the input• Characteristics of the output	instructions, structure, timing, etc. format, topical & language features format, topical & language features
Defining the Construct	❖ Involves many rounds of discussion, design & revision!	
Designing the Specifications	<ul style="list-style-type: none">• Task Design	
Pilot Testing	<ul style="list-style-type: none">• Test Prototype• Finalization	

Designing the Specifications

Reading & Listening Components		Writing & Speaking Components	
Guided by Domain Analysis & Construct	Set by Test Developers	Guided by Domain Analysis & Construct	Set by Test Developers
<u>Features of the text:</u> <ul style="list-style-type: none"> • text type, genre • topics, context • register or formality • language functions • language features 	<u>Features of the text:</u> <ul style="list-style-type: none"> • # of texts • format, length 	<u>Features of the prompts:</u> <ul style="list-style-type: none"> • topics, context • register of formality • language functions • language features 	<u>Features of the prompts:</u> <ul style="list-style-type: none"> • # of tasks • format, length
<u>Features of the items:</u> <ul style="list-style-type: none"> • reading skills • listening skills 	<u>Features of the items:</u> <ul style="list-style-type: none"> • # of items • item format 	<u>Features of response:</u> <ul style="list-style-type: none"> • content • language • organization 	<u>Features of the scale:</u> <ul style="list-style-type: none"> • assessment criteria • scoring methods

Designing the Specifications

Example Specifications	
Testlet Name	Reading Correspondence
Text Type	Personal Email or Letter
Topics	***
Register	***
Language features	***
Language functions	***

Outcome

Task Specifications

- reflect TLU domain & test construct
- meaningful & relevant for test users
- maximize testing potential across target proficiency range

Task Design: Receptive Skills

Reading Task: Reading Correspondence

Hi Mea,

I'm sorry for taking so long to reply to your email. As you know, I've been very busy this summer with Marco's university graduation and my family's visit from Chile. The graduation ceremony was great! Too bad you guys couldn't make it, but we understand it's a bit of a drive from Calgary. You were truly missed. My mom remembers you well from when we were kids. After the celebrations, we took the family sightseeing here in Vancouver. They loved it! They left yesterday for Victoria, and will catch a plane back home from there in three days.

1. Input: reading or listening texts

- same texts used to assess learners across a range of levels

2. Tasks: questions based on texts

- items target different skills to measure range of learner abilities

Task Design: CELPIP Reading

Reading Task: Personal Correspondence

Hi Mea,

I'm sorry for taking so long to reply to your email. As you know, I've been very busy this summer with Marco's university graduation and my family's visit from Chile. The graduation ceremony was great! Too bad you guys couldn't make it, but we understand it's a bit of a drive from Calgary. You were truly missed. My mom remembers you well from when we were kids. After the celebrations, we took the family sightseeing here in Vancouver. They loved it! They left yesterday for Victoria, and will catch a plane back home from there in three days.

Example of a Reading Skills Taxonomy

- Understanding vocabulary & grammar in use
- Identifying main ideas / specific details
- Identifying explicit / implicit meaning
- Understanding textual cohesion
- Synthesizing or integrating information
- Identifying context, purpose, tone, register, etc.

Task Design: Productive Skills

Writing Task: Writing an Email

You recently made reservations for dinner at a very famous and expensive restaurant in town. However, you did not enjoy your meal because the food and the service were terrible.

Write an email to the restaurant's manager in about 150-200 words.

1. Tasks: speaking or writing prompts

- same prompts used to assess learners across range of levels

2. Output: oral or written responses

- assessment scales measure skills across range of levels

Task Design: CELPIP Writing

Writing Task: Writing an Email

You recently had dinner at a very famous and expensive restaurant. You did not enjoy your meal because the food and the service were terrible.

Write an email to the restaurant's manager in about 150-200 words.

Your email should do the following things:

- Describe what was wrong with the food you ordered.
- Complain about the bad service.
- Tell the manager how you want to solve this problem.

Example of Assessment Scale Criteria

- Content/coherence
 - quality of ideas & organization
- Vocabulary
 - range, precision, accuracy
- Readability
 - format, cohesion, grammar, punctuation
- Task Fulfillment
 - relevance, completeness, length

Test Development Process

Domain Analysis	<ul style="list-style-type: none">• The TLU Domain
Defining the Construct	<ul style="list-style-type: none">• Target Contexts• Language Knowledge & Skills
Designing the Specifications	<ul style="list-style-type: none">• Task Design
Pilot Testing	<ul style="list-style-type: none">• Test Prototype• Finalization

Questions

How will the test perform with test takers?

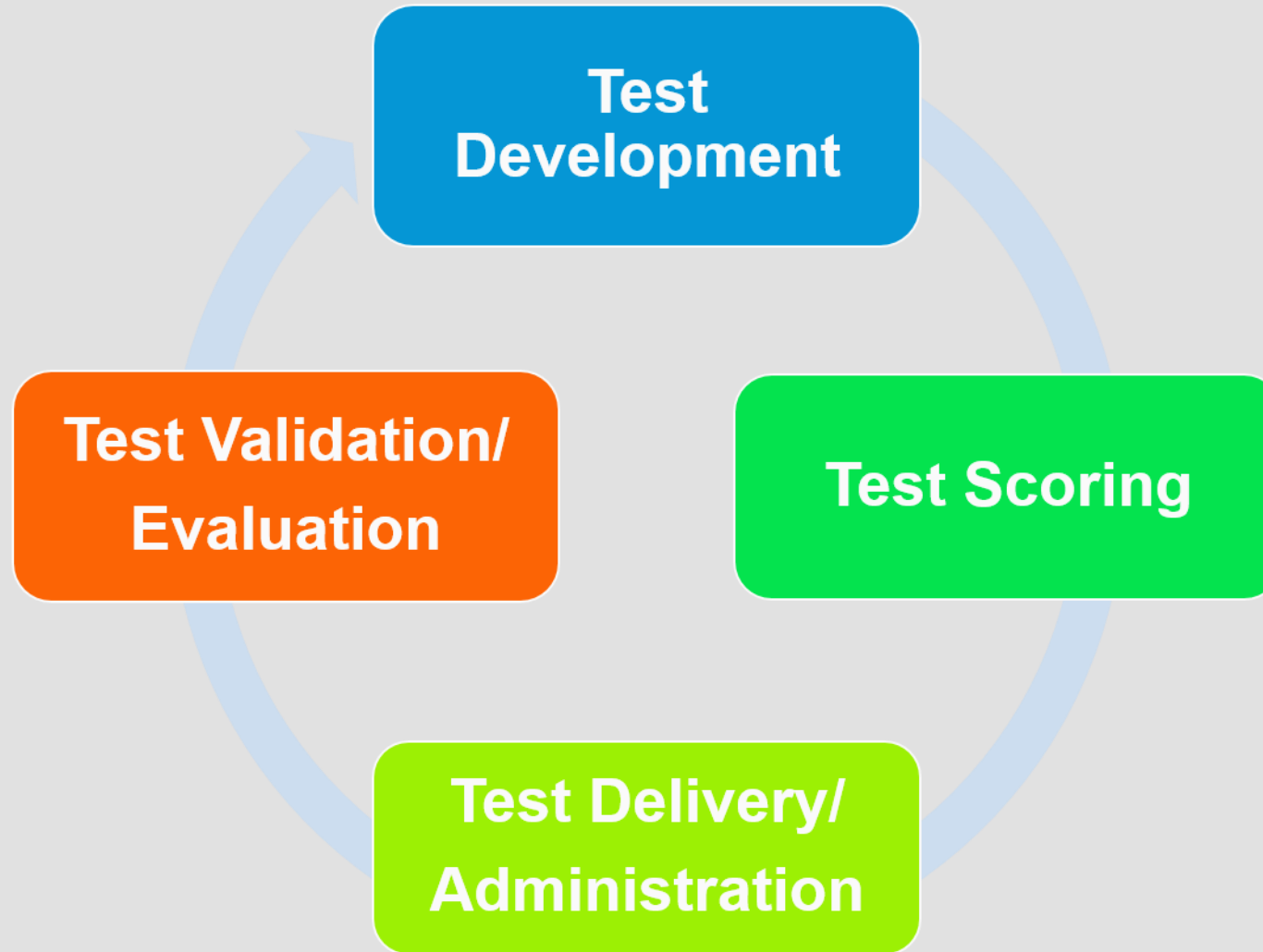
Activities

- Testing the prototype
- Revising & adjusting items

Outcome

Finalizing the test items for administration

Test Development, Operation, and Evaluation Cycle



The CLB as a Standard for Test Alignment

Alignment of Tests during Development

- The CLB as a resource for test development

**Test
Development**

Alignment of Tests by Linking Studies

- The CLB as a resource for linking studies

Test Scoring

Evaluation of Test Alignment by Validation

- The CLB as a resource for test validation

**Test Validation/
Evaluation**

Language Tests

Many types, for different contexts & purposes

- Types/Purposes:
 - Placement – grouping according to ability (specific program)
 - Diagnostic – assessing strengths/weaknesses (instructional objectives)
 - Achievement – determining how much has been learned (course content)
 - Proficiency – determining general language ability



External proficiency standards

Growing interests in aligning tests to external performance standards or frameworks





Why link a standardized proficiency test to the CLB?

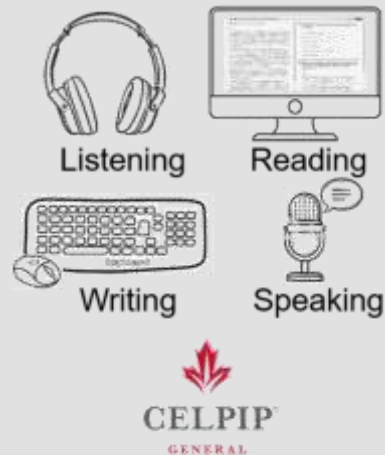
- *Benefits for test takers and score users*

Consistency	Provides a common language for describing proficiency
Transparency	Helps users understand language proficiency requirements
Portability	Facilitates transfer of language proficiency test scores
Interpretability	Expands interpretation of the score beyond the test

Linking to External Standards

Questions

What band level on the test relates to each benchmark?



Linking to External Standards

Activities

- Linking studies to align test results to external standards (CLB)
 - Based on informed expert judgment (panel of experts)
 - Uses psychometric methodology (Angoff Method, Bookmark Method, etc.)
- Reviewing test specifications; making adjustments when necessary



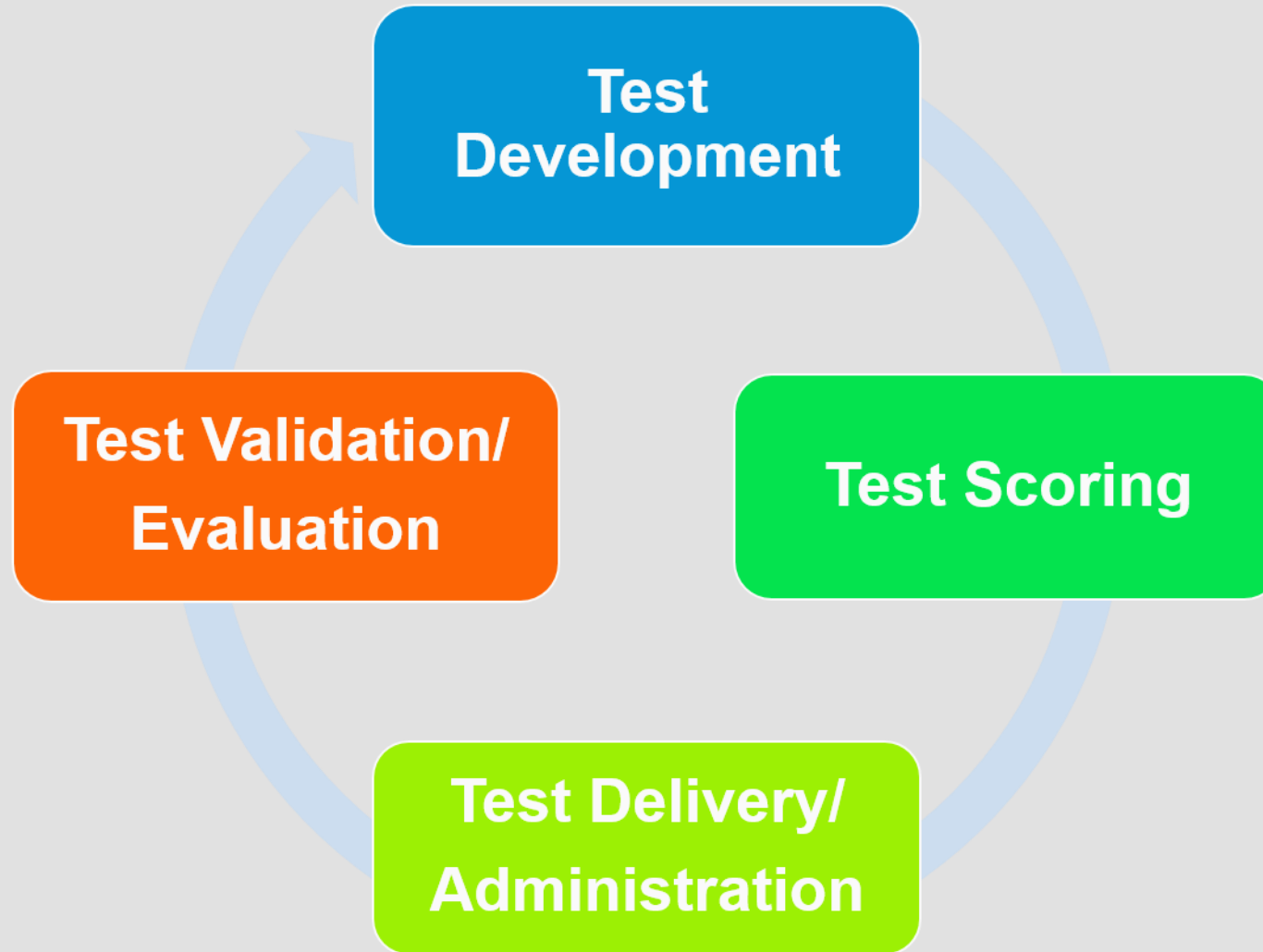
Linking to External Standards

Outcome

1. Establishing a link between CLEPIP Levels & Canadian Language Benchmarks Benchmarks
2. Supporting test score interpretability, one aspect of validity



Test Development, Operation, and Evaluation Cycle



The CLB as a Standard for Test Alignment

Alignment of Tests during Development

- The CLB as a resource for test development

**Test
Development**

Alignment of Tests by Linking Studies

- The CLB as a resource for linking studies

Test Scoring

Evaluation of Test Alignment by Validation

- The CLB as a resource for test validation

**Test Validation/
Evaluation**

Validity Evidence



Test Development

Construct/content
coverage &
representation

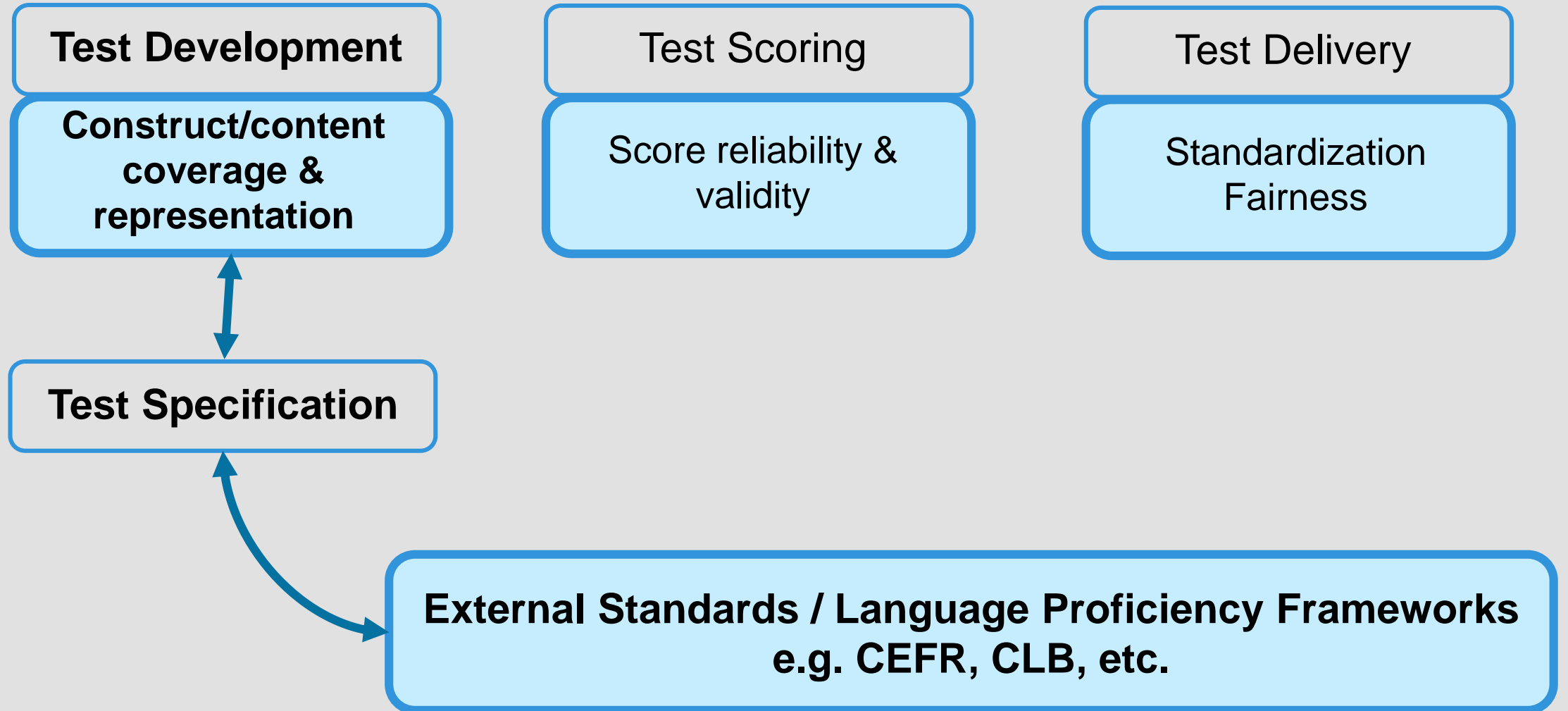
Test Scoring

Score reliability &
validity

Test Delivery

Standardization &
fairness

Validity Evidence for Test Scores Linked to External Standards



Validating the Test Specifications

Questions

To what extent do the tasks reflect the language knowledge & skills described described by the CLB?



Test Validation

How do test developers use the CLB as a resource for test validation?



The CLB as a Resource for Test Validation



Activities

Expert Panel

- mapping features of test content to the CLB descriptors

Evaluation

- coverage of language knowledge & skills across the target proficiency spectrum

Mapping CLB Descriptors to Test Tasks

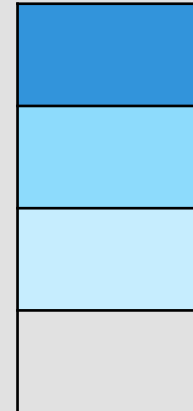
Example. Reading Task (Textual Knowledge)

CLB Stage	Descriptors
Stage 2_1	Understanding cohesive links to interpret contrasts or illustrations by example
Stage 2_2	Understanding paragraph structure and relationships between paragraphs
Stage 2_3	Understanding discourse indicators and patterns signaling cause & effect
Stage 2_4	Understanding textual organization of common written formats in the workplace
Stage 2_5	Understanding an expanding inventory of stylistic devices for narrating & describing
Stage 2_6	Understanding genre, rhetorical forms and their roles in interpretation & coherence

Mapping CLB Descriptors to Test Tasks

Key to correspondence

- + Identified by Paragon & CLB panel
- Identified by members of the CLB panel
- Identified by Paragon developers
- ↓
- Correspondence not found



Mapping CLB Descriptors to Test Tasks

Example: Reading Task (Textual Knowledge)

CLB Stage	Input Text	Item 1	Item 2	Item 3	Item 4	Item 5
Stage 2_1						
Stage 2_2						
Stage 2_3						
Stage 2_4						
Stage 2_5						
Stage 2_6						

Validating Test Specifications

Outcomes

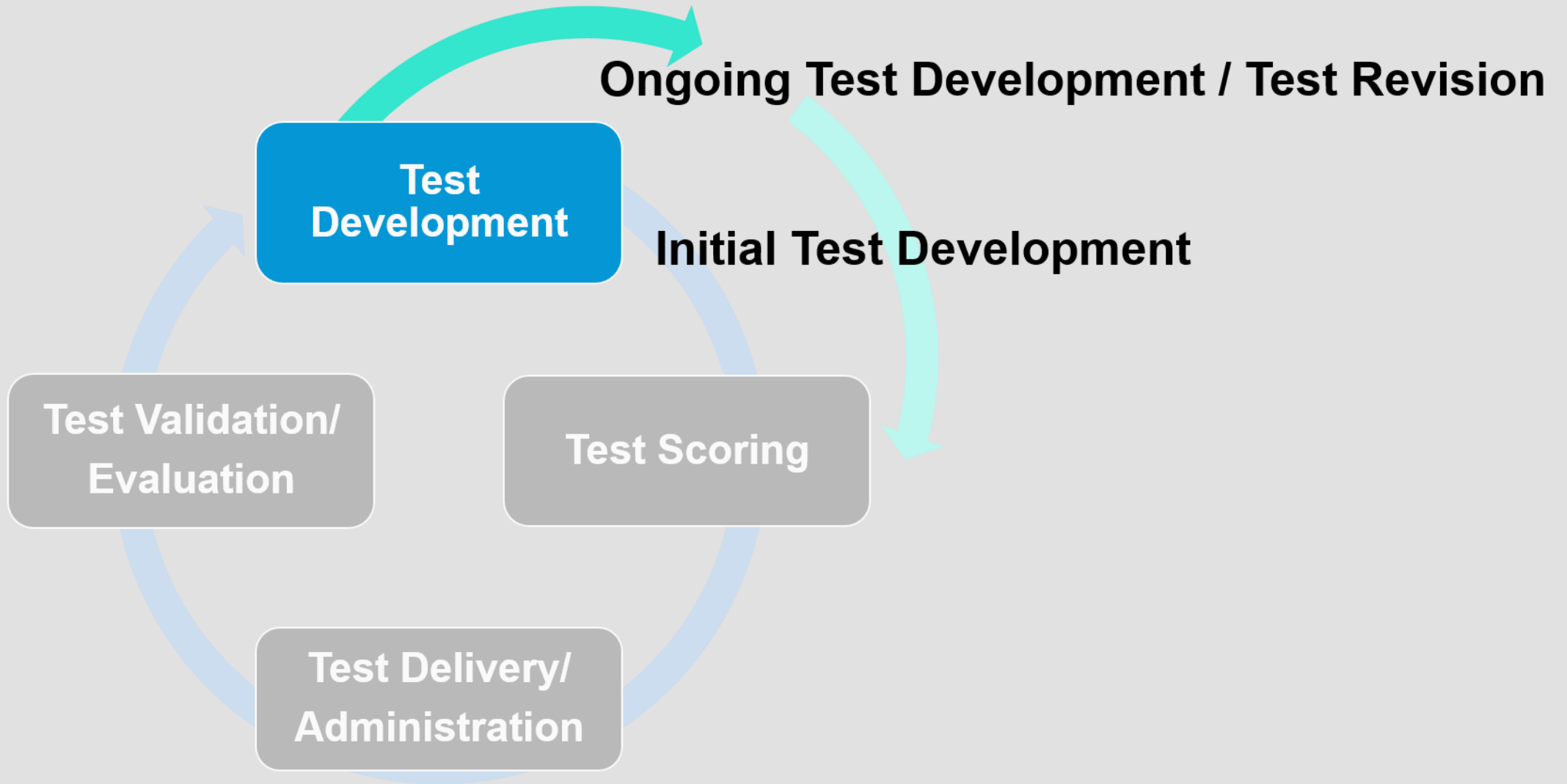
Confirmation of alignment

- Language knowledge & skills required by test compared to CLB standards

Refinement of test specifications

- Filling in the identified gaps as needed





Refining the Specifications

Some considerations when working with external standards ...



- Gaps identifying features described in one level, but not another
- Fairness considering elements not suitable for the testing context
- Terminology understanding the meaning of terms as used by the standard
- Definitions comparing external definitions of features to test definitions

Alderson, Figueras, Kuijper, Nold, Takala & Tardieu (2006)

Gaps

Features may be described in one level, but not in another

Example

CLB Speaking Stage I	'speaks in short sentences'
CLB Speaking Stage II	***
CLB Speaking Stage III	'speaks in connected discourse'

Question for test developers:

- How do we define 'intermediate' speaking (Stage II)?



Fairness

Considering elements that may not be suitable for the testing context

Example

CLB Sociocultural Knowledge: *'use of humour & cultural references'*

Question for test developers:

- To what extent can test takers be expected to have this knowledge?



Terminology

Understanding the meaning of terms as used by the external standard

- Use of terminology

Example 1

CLB contexts: 'demanding' vs 'non-demanding'



Question for test developers:

- How is the concept of 'demand' defined by the CLB?

Terminology (*continued*)

Understanding the meaning of words as used by the external standard

- Use of synonyms

Example 2

CLB Speaking: 'conveys' vs 'expresses' (Stages I – III)

CLB Writing: 'copies' (Stage I) vs 'reproduces' (Stages II & III)



Question for test developers:

- Are they synonyms describing the same underlying ability?

Definition

Comparing definitions used by external standards to test definitions

Example 1

CLB Writing (Stage III): 'lengthy writing' = 1,500+ words, summary of 15+ pages

Question for test developers:

➤ How are features e.g. 'length' defined operationally on the test?



The CLB as a Resource for *Ongoing* Test Development



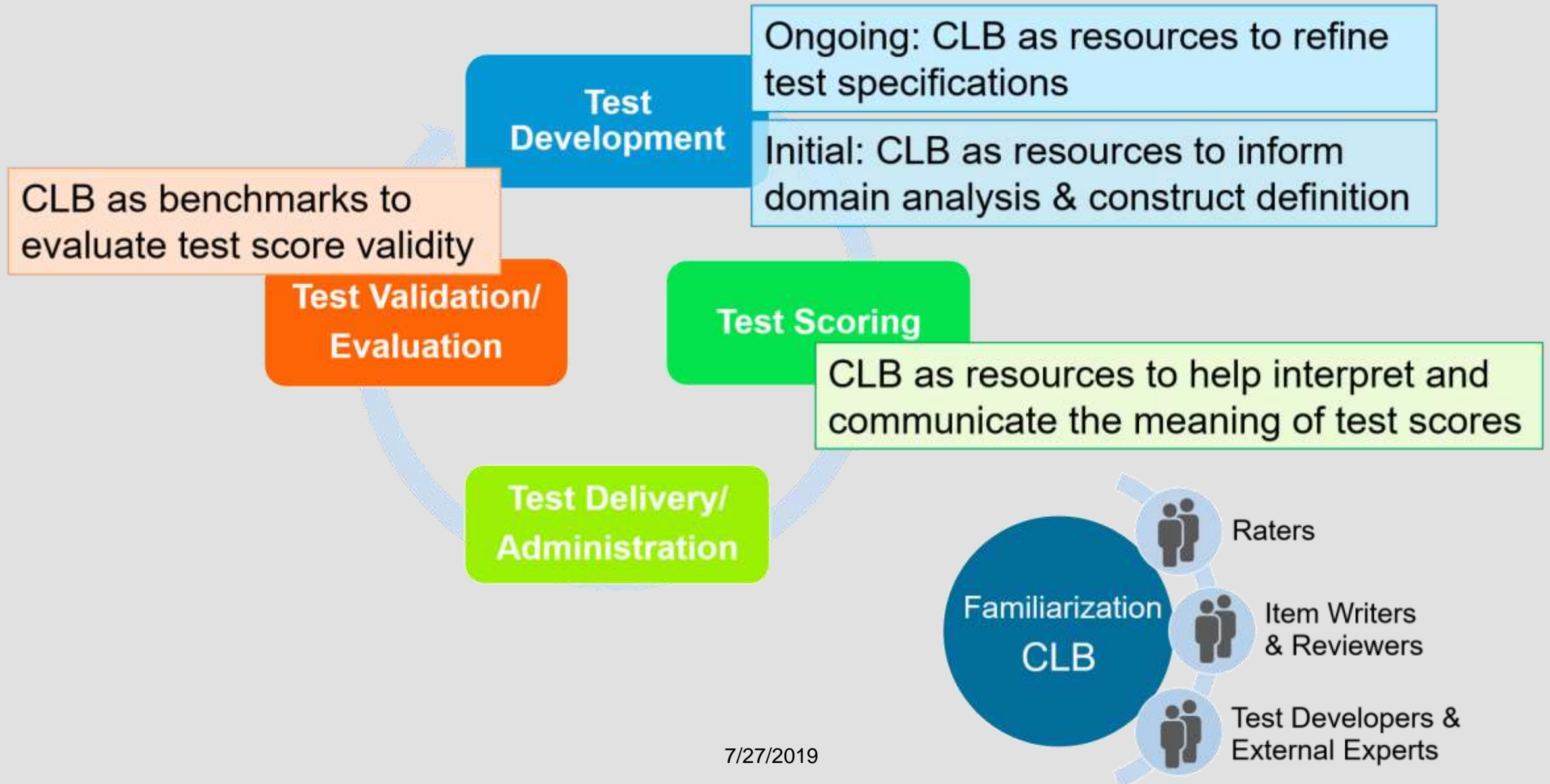
Test Development Team:

- item writers, reviewers & editors
- speaking & writing raters
- Internal test developers & external experts

The CLB as a Resource for Test Development

- For standardization of the language used to describe proficiency
- For consistency in understanding of language learner progression

The CLB as a Resource for Test Development and Validation



Thank You

Contact:

Michelle Chen, mchen@paragontesting.ca

Jennifer Flasko, jflasko@paragontesting.ca

Citations

- Alderson, J., Figueras, N., Kuiper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analyzing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly*, 3(1), 3-30.
- Bournot-Trites, M., Barbour, R., Jezak, M., Stewart, G., & Blouin Carbonneau, D. (2013). *The Theoretical Framework for the Canadian Language Benchmarks and Niveaux de compétence linguistique canadiens*. Ottawa, ON: Centre for Canadian Language Benchmarks.
- Centre for Canadian Language Benchmarks (2012). *Canadian Language Benchmarks: English as a Second Language for Adults*. Retrieved from <https://www.canada.ca/content/dam/ircc/migration/ircc/english/pdf/pub/language-benchmarks.pdf>
- Downing, S. & Haladyna, T. (2006). *Handbook of Test Development*. New Jersey: Lawrence Erlbaum Associates, Ltd.