

Accuracy, False Positive, & False Negative

- Test scores are often used for classification purposes. e.g., pass vs. fail; different grade levels.
- Classification accuracy or decision accuracy (CA/DA) evaluates the degree of accuracy of the classifications.

		TRUE ABILITY LEVEL	
		Pass	Fail
CLASSIFICATION BASED ON TEST SCORES	Pass	True Positive	False Positive
	Fail	False Negative	True Negative

Figure 1. Classification Matrix

Validity and Fairness

- CA/DA is a major concern in criterion-referenced, high-stakes score use because of its close association with validity and fairness.

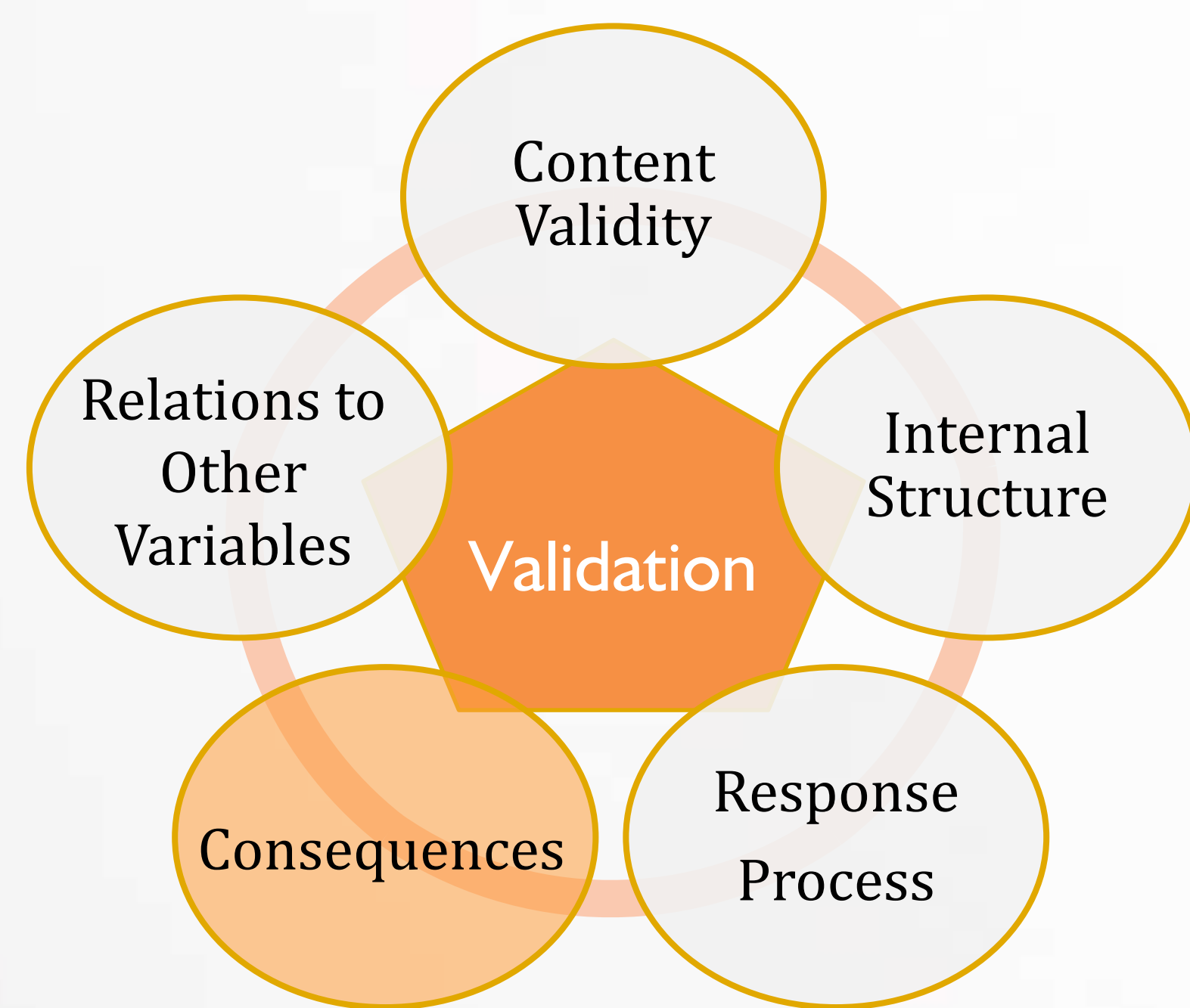


Figure 2. Five Sources of Validity Evidence according to *The Standards* (AERA et al., 2014)

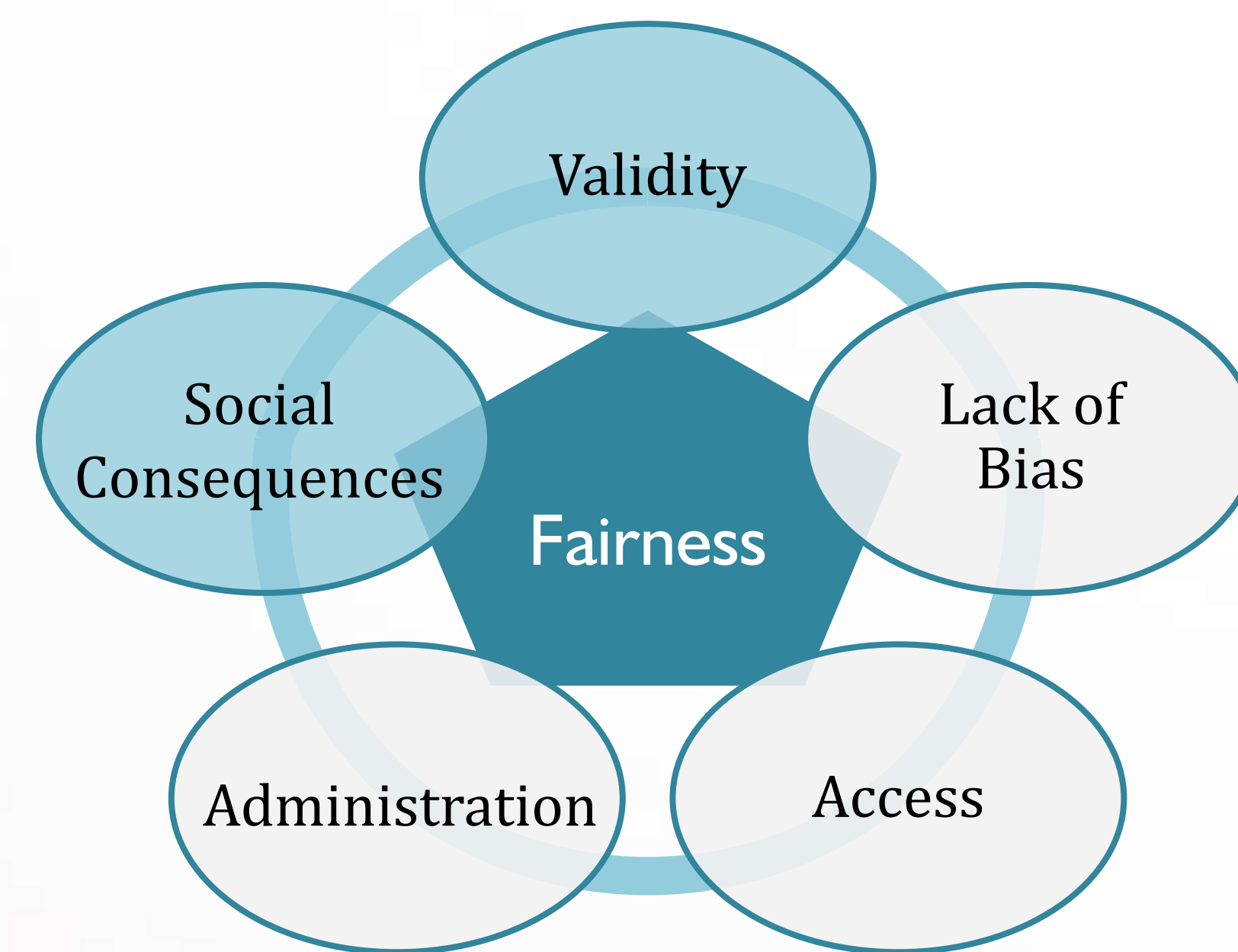


Figure 3. Kunnan's Test Fairness Framework (Kunnan, 2000, 2004, 2008)

Data Source

The Canadian English Language Proficiency Index Program (CELPIP) – General Test measures the functional language proficiency required for successful communication in general Canadian social, educational, and workplace contexts.

Table 1. Critical CLB & CELPIP Levels for IRCC Requirements

CRITICAL LEVEL	PROGRAM AFFECTED		IMPACT
CLB 4	CELPIP Level 4	<ul style="list-style-type: none"> • Federal Skilled Trades: Reading & Writing • Canadian Citizenship: Speaking & Listening 	Minimum requirement
CLB 7	CELPIP Level 7	<ul style="list-style-type: none"> • Federal Skilled Worker • Canadian Experience Class (NOC 0 or A) 	Minimum requirement

Note: IRCC = Immigration, Refugees and Citizenship Canada; CLB = Canadian Language Benchmarks (CCLB, 2012); NOC = National Occupational Classification

Method

- We employed Rudner's method for the listening and reading components (Rudner, 2001, 2005), using the R package cacIRT (Lathrop, 2014).
- We adopted the Livingston and Lewis (1995) method for the speaking and writing components, using the computer program BB-Class (Brennan, 2004).

Results

Table 2. Classification Consistency and Accuracy at Critical Levels

CELPIP LEVEL	COMPONENT	CLASSIFICATION CONSISTENCY	ACCURACY		
			Classification Accuracy	False Negative	False Positive
4	Listening	0.96	0.97	0.02	0.01
	Reading	0.96	0.97	0.02	0.01
	Writing	0.99	0.99	0.00	0.00
	Speaking	0.98	0.98	0.00	0.02
7	Listening	0.90	0.93	0.04	0.03
	Reading	0.90	0.93	0.04	0.03
	Writing	0.88	0.92	0.05	0.03
	Speaking	0.94	0.96	0.03	0.02

Discussion & Conclusion

- The primary purpose of this study was to evaluate consequential validity evidence by assessing the classification accuracy of test takers' language proficiency indicated by a CELPIP score as the level described by the CLB.
- Despite their relevance, classification accuracy and related concepts, such as false positive and false negative, have rarely been used when discussing the validity and fairness of test score use.
- A false negative (i.e., scoring a candidate lower whose true proficiency is at or above CLB 7) denies the candidate and their families fair opportunities;
- A false positive (i.e., scoring a candidate higher whose true proficiency is below CLB 7) results in unfairness to the other eligible candidates, and the need for additional resources to assist the individual with settling in Canada.
- The results show that the CELPIP test scores have high classification accuracy and low false-negative and false-positive rates at the critical levels.