

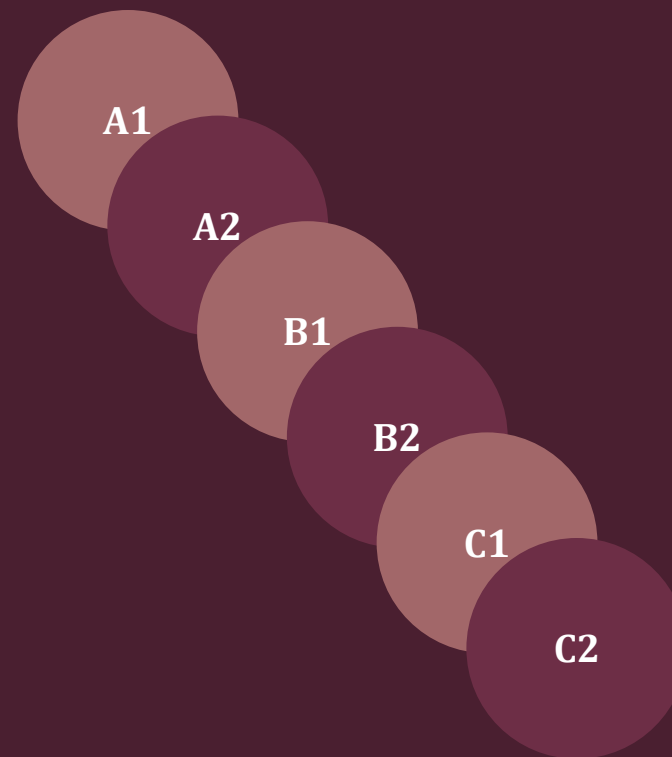
ASIA TEFL 2025

Integrating CEFR in G-TELP

A Study on the Most Effective Assessment Method

Minjung Kim

ITSC GTELP Korea



Two Frameworks, One Question

CEFR sets the global standard for describing proficiency; G-TELP measures it in practice. This study asks how the two should be connected.

C2

CEFR

Common European Framework of Reference for Languages

- Global standard for describing language proficiency (A1–C2)
- Developed by the Council of Europe (2001)
- Six main levels, with intermediate + bands (e.g., B1+)
- Emphasizes communicative competence through task-based descriptors
- Used for curriculum alignment, assessment comparability, and policy

GT

G-TELP

General Tests of English Language Proficiency

- Practical proficiency test evaluating five skills separately
- Listening • Grammar • Reading • Speaking • Writing
- Recognized for criterion-referenced diagnostic feedback
- Growing global interest in aligning G-TELP scores with CEFR levels
- Need to integrate G-TELP into the CEFR framework

CEFR Proficiency Levels at a Glance

A1	Basic	Understands and uses the most familiar, everyday expressions; can introduce themselves if the other person speaks slowly and helps.
A2	Basic	Understands frequently used sentences; can exchange simple information about family, shopping, and surroundings.
B1	Independent	Understands familiar speech and writing from work, school, and leisure; can describe experiences, events, and plans.
B2	Independent	Understands abstract and complex material in one's field; communicates fluently with native speakers on varied topics.
C1	Proficient	Understands long, demanding texts and implicit meaning; uses language fluently and spontaneously for social, academic, and professional purposes.
C2	Proficient	Understands virtually everything heard or read; integrates and restructures information fluently, precisely, and with nuance.

Mapping G-TELP Scores onto CEFR Levels

The challenge: CEFR has 6 main levels (with + bands), while G-TELP tests 5 skills individually — so integration requires a cross-domain evaluation method.

This study compares three conversion methods:

1

Simple Average

Mean CEFR level across all five skills

2

Lowest Score

Final level set by the weakest individual skill

3

Weighted Average

Extra weight given to productive skills
(Speaking/Writing)

CEFR Research & Integration

GLOBAL TREND

- ETS (2008): linked TOEFL iBT and TOEIC to CEFR
- ACTFL (2012): mapped OPIc scores to CEFR
- Cambridge English: aligned IELTS to CEFR (Lim et al., 2013)

SOUTH KOREA

- Lee & Kim (2009): alignment with the Korean national curriculum
- Hwang (2016): CEFR's value for educational efficiency
- Kim (2019): CEFR supports transparency and fairness in testing

GAP IDENTIFIED

- Limited research on multi-domain tests like G-TELP
- No methodologically sound approach yet exists for integrating G-TELP scores into the CEFR scale
- This study addresses that gap by testing and comparing conversion methods directly

Three CEFR Conversion Methods

1

Simple Average

Mean of all domain levels

How it works Mean CEFR level across five skills (L/G/R/S/W)

Trade-off Easy to understand, but lacks sensitivity to weak domains

2

Lowest Score

Conservative, floor-based

How it works Final CEFR level is set by the lowest individual score

Trade-off Ensures minimum competence in all domains, but may be unfair

3

Weighted Average

Emphasizes Speaking/Writing

How it works Higher weight assigned to productive skills (S/W)

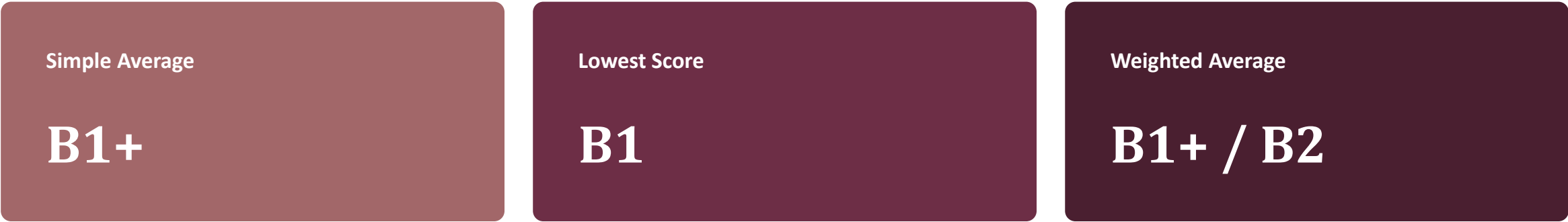
Trade-off Reflects real-life demands in academic and professional settings

Sample Simulation

Simulated candidate profiles with varied domain performance to calculate CEFR conversion across the three methods, and to assess **consistency, fairness, and communicative accuracy**.



RESULTING CEFR LEVEL BY METHOD



Surveys & Expert Interviews

160

General Survey

adults surveyed

- Asked about fairness, usefulness, and preferred conversion method
- Participation incentive provided

13

Expert Interviews

CEFR experts — UK, US, South Korea

- Semi-structured interview format
- Reviewed fairness, clarity, and policy implications

What the Data Showed

EXPERTS

**avored Weighted
Average**

*for balancing domain importance and
fairness*

TEST TAKERS

preferred Simple Average

for its clarity and ease of understanding

BOTH GROUPS

**saw Lowest Score as
overly strict**

*“too penalizing” and demotivating for
uneven skill profiles*

Expert and Survey Perspectives

EXPERT OPINION (n = 13)

- Majority preferred Weighted Average for balancing domain importance
- Several noted it reflects real-life language use more accurately
- Lowest Score described as “too penalizing” and inhibiting fair representation

GENERAL SURVEY (n = 160)

- Simple Average perceived as easiest to understand
- Majority found Lowest Score demotivating
- Weighted Average was favored by students with higher writing/speaking scores

Thematic analysis: transparency in CEFR labeling • weighted conversion allows flexible use across institutions • experts advocate policy-level standardization of CEFR linkage

Methodological Tensions & Implications

1 Fairness vs. Simplicity Test takers prefer clarity (Simple Average); experts favor performance validity (Weighted Average).

2 Equity vs. Rigor The Lowest Score approach protects baseline competencies, but may penalize test takers with uneven skill profiles.

3 Theoretical Grounding The weighted method aligns with the construct validity of communicative language ability (Bachman & Palmer, 2010) and supports domain-specific CEFR reporting.

4 Limitations Further validation is needed using real-world test data from diverse regions and age groups.

IMPLICATIONS

A Staged Path to CEFR-Linked Reporting

1

Phase 1

Use Weighted Average for institutional or academic use

2

Phase 2

Provide detailed domain reports alongside the CEFR band

3

Phase 3

Consider adaptive testing for CEFR-linked score interpretation

APPLICABLE FOR

University admissions

Corporate hiring & global mobility programs

Immigration & government qualification criteria

Enhances G-TELP's global recognition and comparability with other high-stakes tests

CONCLUSION

The Weighted Average Best Balances the Trade-offs

- Among the tested methods, the Weighted Average best balances validity, fairness, and user expectations
- It aligns with CEFR's action-oriented approach and multi-dimensional descriptors
- Adoption is encouraged among policy makers, educators, and test developers

FURTHER RESEARCH

- Cross-cultural validation
- Longitudinal tracking of CEFR-linked performance
- Machine-learning based CEFR prediction models

References

- ACTFL. (2012). ACTFL proficiency guidelines. American Council on the Teaching of Foreign Languages.
- ETS. (2008). Linking TOEIC and TOEIC Speaking scores to CEFR. Educational Testing Service.
- Hwang, H. J. (2016). A study on the application of CEFR in Korean English education. *Foreign Languages Education*, 23(3), 121–144.
- Kim, M. J. (2019). CEFR 적용을 통한 영어교육의 투명성 확보. *Language Assessment Quarterly*, 16(2), 89–104.
- Lee, J. H., & Kim, S. H. (2009). Adapting the CEFR in the Korean national curriculum. *The Journal of Asia TEFL*, 6(4), 123–144.
- Lim, G. S., Lee, Y., & Taylor, L. (2013). Developing speaking assessment criteria aligned with the CEFR. *Cambridge English Research Notes*, 54, 14–20.
- Tannenbaum, R. J., & Wylie, C. (2008). Linking TOEFL iBT scores to the CEFR. *ETS Research Report Series*, 2008(1), i–72.

Thank You

Minjung Kim • ITSC GTELP Korea

A1

A2

B1

B2

C1

C2