資 料 2

「日本語教育の参照枠」二次報告（案）
－日本語能力評価の考え方について－

１．日本語能力評価の現状と課題

○　外国人等の日本語能力を判定する方法として国内外で様々な試験（約 20 の機関・団体）が実施され，個々の指標に基づき，レベルや判定基準等が設定されているが，学習・教育内容の多様化が進む中，各試験が判定する日本語能力についての共通の指標を整備し，利用できるようにすることが必要となっている。

○　例えば，留学生にとって必要な日本語の知識や能力を測る試験で示された日本語能力のレベル判定基準が，そのまま「生活者としての外国人」の日本語能力評価や学習目標の指標として用いられてしまうことがあり，留学生とは異なる目的，場面で言語活動を行う人の日本語能力について，適切な判定がなされていない。

○　現行の日本語能力を判定する試験においては，話すこと，書くことなどの産出に関する能力を評価するものが少なく，かつその評価のための基準も明確に示されているとは言い難い。

○　地域の日本語教室等で日本語を学ぶ学習者の中には，試験による評価を必要としない者も少なくない。そのため，個々の現場で実施できるポートフォリオによる評価などの代替的評価の方法と事例についても　幅広く示していく必要がある。

２．「日本語教育の参照枠」における言語教育観に基づく評価の三つの理念

○　国内外における日本語学習者の日本語の習得段階に応じて求められる日本語教育の内容及び方法を明らかにし，外国人等が適切な日本語教育を受けられ，評価できるようにするため，「日本語教育の参照枠」の考え方に基づき，外国人の日本語能力の判定基準及び評価の在り方について策定する。

○　その際，「ヨーロッパ言語共通参照枠（CEFR：Common European Framework of Reference for Languages: Learning, teaching, assessment ）」（以下，CEFR（2001）という。）を参考に，主に言語知識を測定する筆記試験等による評価だけでなく，パフォーマンス評価やポートフォリオの作成，自己評価などの代替的評価を含めた評価の在り方を示す。

○　国内外で様々な日本語能力を判定する試験が実施され，個々の指標に基づき，レベルや判定基準等が設定されているが，学習・教育内容の多様化が進む中，各試験が判定する日本語能力についての共通の参照枠を整備し，利用できるようにする。

○　「日本語教育の参照枠」では，「日本語教育の推進に関する法律」第一条に掲げる「多様な文化を尊重した活力ある共生社会の実現に資するとともに，諸外国との交流の促進並びに友好関係の維持及び発展に寄与する」ことを理念として示し，言語教育観の柱として以下の三つを示した。

① 日本語学習者を社会的存在として捉える
② 言語を使って「できること」に注目する
③ 多様な日本語使用を尊重する

○ この三つの言語教育観の柱に基づき，「日本語教育の参照枠」における三つの評価の理念を示す。

① 生涯にわたる自律的な学習の促進

「日本語教育の参照枠」における評価は，生涯にわたる自律的な学習の促進[1]を目的とする。

② 学習の目的に応じた多様な評価手法の提示と活用推進

「日本語教育の参照枠」では，日本語を使用して，何が，どのように，どれくらいできるのかを言語能力記述文等を用いて具体的に示すとともに，それがどの程度達成できたかを把握するために，多様な評価手法を提示し，その活用を後押ししていくための考え方や事例を示す。

③ 基準（尺度）と評価手法の透明性の確保

日本語学習者，教師ばかりでなく，一般の日本人等にとっても参照しやすい，日本語で「できること」に注目した基準（尺度）を示し，その評価手法の透明性を確保することを通して，日本語教育に関わる全ての人の間で評価に関する共通認識を醸成する。これにより，日本語学習者がいつ，どこにいても，一貫した学びを継続できる環境の整備を目指す。

1 CEFR（2001）にも「CEFR はさらに学習者の熟達度のレベルを明示的に記述し，それぞれの学習段階で，生涯を通して学習進度が測れるように考えてある。」とある。

３．「日本語教育の参照枠」における言語能力観と評価（何を測るのか）

（１）言語能力観について

CEFR（2001）では，「人間の全ての能力は，言語使用者がコミュニケーションを行う力に何らかの形で寄与することから，あらゆる能力はコミュニケーション能力の一部と考えてよい。それでも，言語とはそれほど緊密に関わらないものを，狭義の言語能力の範疇に含まれるものから区別することは意義あることだろう。」として，言語使用者 / 学習者の言語熟達度を構成する能力及び言語能力を，次の四つに整理して示している。

① 一般的能力

一般的能力とは，叙述的知識（世界・社会文化・異文化などについての知識），技能とノウ・ハウ（生活や余暇・社会的・異文化間・職業的な技能），実存的能力（態度・動機・価値観・信条・認知的スタイル・性格），学習能力（言語とコミュニケーションに関する意識・音声意識と技能・学習技能・発見技能）から構成される。

② コミュニケーション言語能力

コミュニケーション言語能力とは，言語能力，社会言語能力，言語運用能力から構成される。「日本語教育の参照枠」では，これらの能力に基づき能力 Can do を示している。

③ コミュニケーション言語活動

コミュニケーション言語活動とは，受容，産出，やり取り，仲介の四つから構成される。「日本語教育の参照枠」では，活動 Can do として５つの言語活動別（受容的言語活動：「聞くこと」，「読むこと」，産出的言語活動：「話すこと（発表）」，「書くこと」，相互行為的言語活動：「話すこと（やり取り）」）の言語能力記述文を示している。

④ 方略

方略とは，言語活動を行う上で駆使する，分からない言葉などに対する推測や質問，聞き取りにくい言葉について聞き返したりする行動を指す。

（２）評価について

CEFR（2001）では，評価の議論には，伝統的に基本となる以下の三つの概念があるとしている。

・妥当性（validity）
　あるテストや評価の手法が，当該の状況で，実際に測っているものと，測っているはずのものとが一致しており，またそこで集められた情報が当該の学習者の熟達度を正しく示している場合に，そのテストや評価には妥当性があるということができる。

・信頼性（reliability）
　基本的に同じ評価が（実際に，または仮定の上で）二回実施された場合，そこで評価された学習者の序列が同じになるかどうかの蓋然性。

・実行可能性（feasibility）
　その評価の手法が現実的に実行可能であるかどうか。

CEFR（2001）では，評価についての論点として，「評価の方法や伝統はさまざまであるが，あるアプローチ（例：教師による評価）より，別のアプローチ（例：公的な試験）の方が，教育上の効果において絶対に優れていると考えるのは間違いである。本書にある共通参照レベルのような，一連の共通基準の主要な利点は，まさにお互いに異なる評価の形式でも対応付けが可能になることである。」ことを挙げ，教育の目的に応じてさまざまな手法を組み合わせて評価を行うことを推奨している。

※参考：CEFR における評価

【第九章　評価(Assessment)】(CEFR　2001，吉島・大橋訳 2014 p.199)

・CEFR では，評価を言語学習のプログラムの総括・広義の評価(evaluation)[2]
という広い問題ではなく，限定的な意味でのアセスメント(assessment)として扱っている。

総括・広義の評価（evaluation）
- ・学習者の熟達度についての評価(assessment)
- ・ある方法や言語教材の効率性
- ・言語学習のプログラムで実際に産出されたディスコースの種類や質
- ・学習者/教師の満足度
- ・教育の効率性

2　Evaluation の訳語として「総括」は一般的であるとは言えず，また，10 ページに出てくる「総括的評価（Summative assessment)」との混乱を避けるため，言葉を補った。

４．「日本語教育の参照枠」における多様な評価の在り方と事例（どう測るのか）

CEFR（2001）では評価を「学習者の熟達度についての評価（assessment）」として扱っていることから，本報告でも，評価については熟達度評価を中心に示すこととする。

（1） 主な熟達度評価の在り方

① 試験として実施することができる評価

主な評価の在り方として，試験として実施することができる評価には，筆記試験とパフォーマンス評価がある。筆記試験については，近年，従来型の紙ベースの試験のほかに，CBT（Computer-Based Testing）が導入されており，これも広義の筆記試験に含むものとする。

○ 筆記試験

ある教育プログラムにおける試験では，ある期間内に扱った学習目標の到達度を測る試験と，その時点で何ができるかという熟達度を測る試験がある。熟達度を測る試験については試験団体が実施する試験を受ける場合もある。

- 学習対象の学習成果を表わす熟達度を，より一般的な言語能力尺度上に位置づけて表わすことができる。
- 試験内容を特定，開発する際に活用できる。
- 異なるテスト間の測定道具としての共通性および違いを明確にできる。
- 国や機関を超えて共通に参照できる，日本語能力を評価する枠組みや構成概念の設定及び，測定道具（テスト）の仕様を検討する基本設計図として活用できる。
- 学習者が自身で言語能力の目標設定や評価（到達点の確認，調整）についての見通しを持つことができる。

○ パフォーマンス評価

パフォーマンス評価とは，産出活動（「話すこと（やり取り）」，「話すこと（発表）」，「書くこと」）についての熟達度を測ることを指す。

パフォーマンス評価は到達度，あるいは熟達度を測る試験として実施する場合と，例えばポートフォリオ作成などの試験によらない代替的な評価として実施する場合がある。就労場面における分野別の産出活動についての熟達度を適切に測ることについては，社会的に高いニーズがある。

・教師と学習者の双方がパフォーマンスに関する評価指標を共有することで，評価の透明性を高めることができる。

・学習者は与えられたパフォーマンス課題に対して，評価指標を基にした明確なフィードバックを得ることができる。

② 試験によらない評価（代替的評価）

○ ポートフォリオによる評価

ポートフォリオによる評価とは，多様な広がりを見せる学習者の達成成果を学習者の様々な必要性，性質や資質に応じて記述し，評価することである。CEFR（2001）では，複文化・複言語能力の育成を推進するため，European Language Portfolio（ELP）[3]を用いて学習者一人一人がさまざまな面から自分の言語発達を記録できるようになっている。

・学習者自身，教師，学習者の周りの人が，学習者の言語の熟達度の成長の過程を通時的に把握することができる。

・学習者が学習機関を移動し，新たに学習を始める際にポートフォリオを示すことで，教師はその学習者がこれまで学んできた内容と熟達度が把握でき，適切な教育内容を準備することができる。

○ 自己評価

自己評価とは，言語能力記述文のリストで構成された自己評価表などを用いて，自分の言語熟達度を把握することの他，学習に対する振り返りを記述することを通して，自律的な学習能力を育成する。

・試験による評価と同様に，学習者が自身で言語能力の目標設定や評価（到達点の確認，調整）についての見通しを持つことができ，学習に対する意識を高めることができる。

○ 相互（ピア）評価

相互（ピア）評価とは，学習者とその周りの人が相互に評価を行うことである。クラスメートや家族，就労場面であれば職場の人などから，日本語の使用についてのコメントを得ることで，自分の熟達度を多面的に把握することができる。

---

[3] https://www.coe.int/en/web/portfolio（令和２年８月 20 日閲覧）

・クラスメートや周りの人々から学習についてフィードバックを得ることで，学習に対する動機を高めることができる。また，他の学習者の評価に関わることを通して自己評価に対する内省を深めることができる。

○　ルーブリック評価

ルーブリック評価とは，例えばやり取り能力の熟達度をパフォーマンス評価によって測る場合に，言語的課題（例えば「家族を紹介する」などの言語的タスク）の達成度に加え，文法的正確さや使用語彙の範囲，発音などの質的側面等の観点を組み合わせた評価表を作成し評価を実施することを言う。また，ルーブリック評価は，点数によって測ることが難しいポートフォリオにおける振り返り記述の内容等の評価にも用いられる。

・パフォーマンス評価については，単にできた / できない，だけの評価だけでなく，何が，どのくらいできたのか，について，多様な観点から評価を行うことができる。

・ポートフォリオにおける振り返り記述などについて，記述内容についての評価の観点を盛り込むことで，質的な評価が可能となる。また，評価の各段階に点を与えることによって，数値化が難しい評価対象に点数を与えることができる。

（２）評価の種類

CEFR（2001）では，評価の種類について以下の項目を挙げている。表１は網羅的なものではなく，ある用語が右に置かれるか，左に置かれるのかも重要ではないとしている。

表１　評価の種類[4]

1	到達度評価 Achievement assessment	熟達度評価 Proficiency assessment
2	基準準拠型評価 (NR) Norm-referencing	規準準拠型評価 (CR)Criterion referencing
3	合否型基準準拠型評価 Mastery leaning CR	連続型基準準拠型評価 Continuum CR
4	継続的評価 Continuous assessment	定点評価 Fixed point assessment
5	形成的評価 Formative assessment	総括的評価 Summative assessment
6	直接評価 Direct assessment	間接評価 Indirect assessment
7	運用評価 Performance assessment	知識評価 Knowledge assessment
8	主観的評価 Subjective assessment	客観的評価 Objective assessment
9	チェックリスト評定 Checklist rating	パフォーマンス評定 Performance rating
10	印象評価 Impression judgment	指針に基づいた判断 Guided judgment
11	全体的評価 Holistic assessment	分析的評価 Analytic assessment
12	シリーズ評価 Series assessment	分野別評価 Category assessment
13	他者評価 Assessment by others	自己評価 Self assessment

1 到達度評価 / 熟達度評価

到達度評価は，特定の目的の達成の度合いを評価し，学習したことを評価する。それゆえ，当該の週や学期に行った 勉強，教科書，シラバスに関連する。到達度評価は各々の授業に基づいた，内部の見方を反映している。つまり，限られた学習範囲の目標（試験範囲）にどの程度到達したかを見るということである。

熟達度評価は，実世界の問題に対して，学習者が，何ができるか何を知っているかの評価である。これは外部からの見方を反映している。

教師は教育へのフィードバックを得ようとして到達度評価の方に自然に関心を向けがちである。雇用者，教育行政の管理者，大人の学習者は，熟達度評価 ，つまり，成果や何ができるようになったかの方に，より関心があるだろう。到達度評価の利点は学習者の経験との差が少ないことである。熟達度評価の利点は誰でもその学習者のいる位置が分かることであり，結果が明確なことである。

---

[4] 表の翻訳と各項目の説明は，吉島茂・大橋理枝 訳・編（2014）pp.205-214 を抜粋し，一部修正した。

2 基準準拠型評価 / 規準準拠型評価

基準準拠とは，学習者に序列をつけ，一緒に学習している他の学習者との相対的な位置を明らかにする評価である。規準準拠は，周りの学習者の力量とは無関係に，その教科の学習者本人の力量だけを純粋に評価するもので，基準準拠に対置されるものでもある。基準準拠はクラスという範囲の中で行うこと（あなたはクラスで 18 番目です）も，人口統計的な同類集団（あなたは 21,567 番目です，上位 14%にいます）など，あるテストを受けた学習者グループの範囲の中で行われることもある。

規準準拠は，個人個人のテスト結果が規準表全体のどこに位置しているか分かるようにするもので，熟達度を垂直軸に，関連領域を水平軸にとって図示するものである。これには（ａ）それぞれのテスト / モジュールがカバーしている関連領域の定義，および（ｂ）「区切り点」，つまり特定の熟達度 の水準に達していると認定できるテストの点の特定が必要となる。

3 合否型基準準拠型評価 / 連続型基準準拠型評価

合否型基準準拠は，単一の「最低限の能力の標準」や「区切り点」を決めて，学習者を「合格者」「不合格者」に分ける仕組みである。このやり方では，学習目標達成の度合いは問わない。

連続型基準準拠は，当該の分野において，個々の力量が予め決められた連続体の中のどの位置にあるかを示すものである。

CEFR は合否型でも連続型でも利用できる。連続型で用いられたレベルの尺度　は共通参照レベルに照合することができる。合否型で設定される目標は CEFR の提案したカテゴリーとレベルの概念表の中に位置づけることができる。

4 継続的評価 / 定点評価

継続的評価は教師，またある場合には学習者から見た授業コース全体を通しての授業中の言語運用，課題やプロジェクトの評価である。最終的な成績は授業コース / 学年 / 学期全体を反映する。

定点評価とは，ある特定の日，すなわち，普通は授業コースの最後か開始以前に行われる試験やその他の評価に基づいて，成績が与えられ判断が下されることである。以前にあったことは問題ではなく，その人が今できることが決定的に重要なのである。

評価は，何らかの判断を下すために一定の点でとり行われる授業コースの外のものであるとみなされることが多い。継続的評価は，授業コースの中に組み込まれており，その授業の終了時に何らかの総合的なやり方で評価する時に使われる。宿題や，学習の強化のための定期的達成度テストとは別に，継続的評価は教師や学習者によるチェックリストや表の形をとることもある。

5 形成的評価 / 総括的評価

形成的評価は，学習の進み具合や学習者の強み，弱点に関する情報を集める継続的な評価である。教師はこれらの情報を授業コースの計画や学習者へのフィードバックに役立てることができる。形成的評価という言葉は，広い意味で用いられることが多く，質問紙や話し合いから得られた数量化できない情報も含まれる。

総括的評価は授業コースの終わりにこれまでの成果を成績としてまとめるものである。それは必ずしも熟達度の評価ではない。事実，総括的評価の多くは標準準拠型の定点評価であり，達成度評価である。

6 直接評価 / 間接評価

直接評価は，学習者が実際にしていることを評価することである。例えば，小グループで何かを論じているところで，評価者がそれを観察して，基準となる表と比較し，言語運用を表の中の最も適当なカテゴリーと一致させ評価を下す。間接評価は通常紙面テストを用い，実行可能だと考えられる技能を評価する。

直接評価は事実上，話すこと，書くこと，やり取りでの聞くことに限られる。というのも，受容的活動は直接見ることができないからである。例えば，読むことについては，学習者に適当な解答欄をチェックさせる，文を完成させる，質問に答えさせるなどして，理解の証を出させることで，間接的に評価するしかない。言語の使用の幅とその把握の程度は，基準との一致度を判断して直接的にも評価できるし，あるいはテスト問題の答えを解釈し，一般化することで間接的にも評価できる。古典的な直接テストは面接であり，古典的な間接テストはクローズ・テスト（cloze test）である。

7 運用評価 / 知識評価

運用評価を行うためには，学習者が実際の発話か，書いた文書の実例を提示しなければならないが，それらは直接テストによって得られる。

知識評価では，学習者がさまざまな種類の質問に答えることになるが，その質問は，学習者がどの程度言語的な知識を持っており，その使い方をどの程度把握しているかを証明するものでなければならない。

能力を直接に測ることはできない。それを測ろうとする場合，運用の幅から熟達度についての一般化を行うしかない。熟達度というのは，実際に使用された能力と考えてよいだろう。この意味であらゆるテストは，運用例を証拠としてその根底にある能力を推定しようとするのだが，全てのテストが測っているのは実際には運用のみである。

8 主観的評価 / 客観的評価

主観的評価は，評価者の判断によって決められる。普通，これは運用の質に対する判断のことである。

客観的評価は，主観性を排した評価である。普通，これは例えば選択肢式のテストのように，各項目に該当する正解が一つしかないような間接テストを意味する。

しかし，主観性 / 客観性の問題は，これよりはるかに複雑である。間接テストは，しばしば「客観式テスト」であるといわれているが，これは採点者が確定的な正解をもとに，受験者の回答を正しいとするか誤りとするかを決め，正しいとした回答の数を数えて最終的な結果を出す類のものを指している。この過程をもう一段階先に進め，それぞれの質問に正解が一つだけしかないようにし，採点者の誤りを防ぐために機械で採点することもしばしばである。実際，この意味での「客観式テスト」の客観性は，多少強調され過ぎているきらいがある。というのは，誰かが，評価という行為をより制御しやすいテストの実施技術に還元するということを決定したからである。そして，誰かがテストの細目を書き，他の誰かが，その細目の中の特定の項目を実際に測定可能な形に書いたのかもしれない。結局は，誰かが，出題される可能性のある他の項目ではなく，その特定の項目をこのテストのために選んだのである。これらの決定は全て何らかの主観的な判断を伴うので，このような形式のテストは客観採点式テストと呼んだ方がよいのではないだろうか。

直接運用評価では，たいてい評価者の判断を基に成績がつけられる。すなわち，学習者がどの程度上手に言語を運用したかという判断は主観的に行われるのであり，関連する諸事情を考慮に入れたり，ガイドラインや基準や経験に照らして決められる。言語やコミュニケーションは非常に複雑であり，自動化に馴染まず全体が個々の部分の総和よりも大きいのだから，主観的な方法には利点がある。特定のテスト項目が実際には何をテストしているのかを明確にすることは困難な場合が多い。従って，能力や運用力の特定の面に焦点を当てたテスト項目というのは，表面的にはともかく，実際はそううまくは機能しない。

しかし，公平を期すためには，すべての評価はできる限り客観的でなければならない。内容の選択や言語運用の質に関する主観的な決定に，個人的な価値判断が影響することは可能な限り減らさなければならないし，特に総括的評価が行われる場合には尚更である。というのは，テストの結果は，その評価を受けた人の将来を決めるために第三者によって使われることが多いからである。

9 チェックリスト査定 / 尺度評定

尺度査定：いくつかのレベル，あるいはレベル帯から構成される尺度に基づいて，学習者が特定のレベルにある，または特定のレベル帯の範囲内にいることを判断する。

チェックリスト査定：特定のレベルやモジュールに関連があるとされる事項のリストに照らして，学習者に対する判断を下す。

「尺度査定」では，学習者をいくつかあるレベル帯のどれかに当てはめることに主眼が置かれる。強調されるのは垂直性であり，学習者が尺度内のどの程度上の位置まできたのか，という点である。それぞれのレベルやレベル帯が意味するものは，その尺度の言語能

力記述文によって明らかにされていなければならない。さまざまな分野に対して，それぞれ複数の尺度があるかもしれないし，それらは表として同じページに記載されていてもよいし，別のページに記載されていてもよい。それぞれのレベルやレベル帯について定義があるかもしれないし，一つおきにしか定義されていなくてもよい。もしくは，上，中，下のレベルに当たるものにしか定義がないこともありうる。

これに代わるものとして，チェックリストがある。これはそのリストに出ている項目に関連する分野を達成できたかどうかをチェックすることが主眼となる。つまり，水平性が強調されるのであり，そのモジュールの内容をどのくらい達成することができたかが重要になる。チェックリストは，質問紙のように，要点を列挙したような形で示すこともできる。一方，それは，車輪のような形で示すこともできるし，また別の形で示すことも可能である。答え方は，「はい / いいえ」だけかもしれないし，もっと細かい形（例えば「０から４」の段階付けで示されるなど）かもしれないが，その場合それぞれに表示がきちんと付いた目盛りがあり，その表示も定義されていることが望ましい。

言語能力記述文は，独立した，当該のレベルと対応した基準項目になっているので，これを基にして特定のレベルのチェックリストを作ることも，全てのレベルに関する査定尺度や表を作ることも両方可能である。

10 印象評価 / 指針に基づいた判断

印象：学習者の授業中の言語運用の経験に基づいて行われる完全に主観的な判断で，特定の評価に関して何の基準にも基づかない評価。

指針に基づいた判断：何らかの基準に基づいて，意図的に評価しようという意識を持って判断を行うことによって，印象のみによる判断を補い評価者の主観性が減じられる判断。

「印象」とは，教師や学習者が，授業中や宿題などの出来具合のみを基にすることによって，判断を下す場合のことを指している。主観的な査定，特に，継続的な評価で使われる査定は，反省や記憶を基にしている場合が多いが，その際に使われている反省や記憶の焦点は，対象者を一定の期間意識的に観察することによって定まってくる可能性が高い。非常に多くの学校でこの方法が実施されている。

「指針に基づいた判断」とは，上に述べたような印象判断が，一連の評価方法を通じて，判断が熟慮に基づいたものになった場合を指している。そのような方法は，(a)何らかの手順に従った評価が行われていること，及び/または(b)各評 点または成績の間を区別できる明確な基準を設けていること，そして(c)標準化のために評価者が何らかの訓練を受けることを意味している。

指針に基づいて判断を行う利点は，このようにして評価する人たちの中で共通基準が確立できれば，下される判断の一貫性が劇的に増すことである。もし，言語運用の実例や他の評価方法との関連が固定的「水準点」として示されているならば，一貫性はさらに向上する。この点の重要性が強調されるのは，さまざまな学問分野で，次のような研究結果が重ねて確認されていることによる。すなわち，評価者が充分訓練されていないまま判断を

下した場合，評価者の厳しさの差が，学習者の実際の能力差と同じくらい大きくなり，学習者に対する評価結果がほとんど偶然で決まったも同然になりかねないということである。

共通参照レベルの尺度の言語能力記述文は，上の(b)のような明確化された基準を示すのに利用できるし，既存の基準によって表示されている標準値が，共通レベルのどの辺りに相当するのかを位置づけるのに使うこともできる。

11 全体的評価 / 分析的評価

全体的評価というのは，包括的で統合的な判断を下すことである。さまざまな評価側面の比重は評価者の直感によって定められる。分析的評価は評価側面の一つ一つを別個に見る。この区別には二通りの仕方がある。すなわち，（a）何を評価するか，（b）どのようにしてレベル帯や級や得点が与えられるかである。ある部分では分析的な評価を行い，別の部分では全体的な評価を行うというように，組み合わせて評価が行われるような方法が採られることもある。

（a）評価対象：

「話すこと」や「言葉のやり取り」のように，包括的な分野に対して何か一つだけ得点や級を与えるように評価しようとする場合もある。他の，もっと分析的なやり方の中には，学習者の言語運用の中で，相互に独立したいくつかの側面ごとに別々に結果を出すことを評価者に要求するやり方もある。さらに，評価者が学習者に対する包括的な印象を記し，それぞれの分野別に分析的な評価を行い，その後に熟慮して全体的な判断を下すようなやり方もある。

（b）結果の算定：

学習者の言語運用を観察して，全体的な観点から尺度の言語能力記述文に当てはめるというやり方があるが，この場合尺度が全体的なもの（包括的な尺度を一つだけ用いるとき）である場合も，分析的なもの（3〜6分野に分かれて表になったもの）である場合もある。このようなやり方は，結果を算定するのに計算を用いない。結果は一つの数値で表されるか，複数の分野にわたって「電話番号」型に羅列したものとなる。他の，さらに分析的なやり方としては，分野別に何らかの評点を与え，それらを合計してその学習者の評価値とし，さらに場合によってはその評価値を成績に変換するというやり方がある。この方法の場合，分野別に配点比重を変えて計算するというのが典型的なやり方である。つまり，さまざまな分野がそれぞれ同等の価値を持っているとは見なされないということである。

12 シリーズ評価 / 分野別評価

分野別評価は，単独の評価課題で学習者の言語運用を評価基準表に照らして判断するやり方である。

シリーズ評価は，相互に関連性のない複数の評価課題を行い，それを一つの全体的な査定結果として，それぞれの段階の内容が示された尺度，例えば0～3や1～4などで，表現するものである（この場合の評価課題は，他の学習者や教師とのロールプレイという形を取ることが多い）。

シリーズ評価は，分野別評価では一つの分野の評価が他の分野の評価に影響するという傾向に対処する方法の一つである。初級レベルでは課題の達成に重点が置かれ，その学習者は何ができるのかということを，単なる印象からではなく実際の言語運用から教師・学習者が評価したものを基にして，チェックリストを埋めていくことが意図される。上級レベルでは，言語運用の中の特定の側面の熟達度を示すような課題が与えられるであろう。結果はその学習者の輪郭像として報告される。

13 他者評価 / 自己評価

他人による評価：教師または評価者による判断

自己評価：自分自身の熟達度の判断

上で述べた評価技術の多くには，学習者自身が関与することも可能である。今までになされた研究から，（例えば，ある授業コースに入れてもらえるかどうかなどの）「高い賭け金」がかかっていない限り，テストや教師による評価は自己評価によって効果的に補完できるといわれている。自己評価の正確さが増すのは以下の場合である。

(a) 明確な基準をもった熟達度の言語能力記述文に基づいて評価が行われる。

及び / または,

(b) 評価が具体的経験と関連する。

この経験そのものがテスト課題であってもよい。また，学習者が評価を行うために訓練を受ければ，自己評価はさらに正確なものとなるだろう。このように系統立てて行われる自己評価と教師による評価やテストとの相関係数（一致の妥当性のレベルの指標）は，教師が行った評価と評価の間や，テストとテストの間や，教師による評価とテストとの間で通常みられる相関係数と同じくらいに高いこともある。

しかし，自己評価の最大の可能性は，それを学習者の動機付けや意識を高めることに使うことにある。学習者が自分の長所に気づき，弱点を認識し，学習の方向付けをさらに効果的なものにする手助けをすることである。

（３）筆記試験によらない評価の事例

筆記試験によらない評価は，代替的評価とも呼ばれ，試験では測ることが難しいとされる，言語を用いた課題遂行能力や学習過程における様々な気付きや学びを把握するための評価の方法のことを指す。言語能力の熟達度の評価は，コースで設定した学習目標や学習者の特性に応じて，試験と試験によらない評価を組み合わせて総合的に実施していくことが望ましい。以下は，主な代替的評価についての事例である。

① パフォーマンス評価の例

・ACTFL-OPI（The American Council on the Teaching of Foreign Languages-oral proficiency interview）は，汎言語的に使える会話能力テストであり，初級から超級までの 10 レベルがある。

・国際交流基金（2016）では，JF 日本語教育スタンダードに準拠したロールプレイテスト（A1～C1）を公開している。

② ポートフォリオ評価の例

・文化庁国語課（2012）は，「生活者としての外国人」向けの「日本語学習ポートフォリオ」を公開し，ポートフォリオによる評価の方法を示している。

・地域の日本語教室では，吹田市国際交流協会（子供ポートフォリオ「学習ノート」），磐田国際交流協会（日本語学習ポートフォリオ）がポートフォリオを作成している。

・国際交流基金の JF 日本語教育スタンダードでは，ポートフォリオを評価の中心的な柱の一つとして位置づけ，世界各地でそれぞれの現場独自のポートフォリオを作成し，評価活動を行っている。

③ 自己評価の例，相互（ピア）評価の例

・CEFR スイスプロジェクトの自己評価チェック表（Self-assessment checklist）[5] では，各レベルの言語能力記述文についての評価を，自分で行うばかりでなく，教師や他者にしてもらう欄が設けられている。また，個々の言語能力記述文が自分の目標かどうかのチェック欄があり，学習者にその項目の重要性も考えさせることができる。この項目によって何をいつ学ぶかについて教師が一方的に決めるのではなく，学習者が自ら判断することによって自律的な学習を促す効果がある。

---

[5] https://www.laits.utexas.edu/fi/sites/laits.utexas.edu.fi/files/Self%20Assessment%20Checklist%20European.pdf（令和 2 年 10 月 30 日閲覧）

５．　日本語能力の判定試験と「日本語教育の参照枠」の対応関係を示す方法

（１）日本語能力の判定試験と「日本語教育の参照枠」の対応関係を示すことの意味

現在，国内外で実施されている日本語能力の判定試験（約 20 の機関・団体）は，個々の指標に基づき，レベルや判定基準等が設定されている。これらの試験が「日本語教育の参照枠」との対応付けを行うことによって，共通の指標での評価が可能となり異なる試験間の通用性が高まる。また，共通の指標での評価を得ることができることで，受験者はどの試験を受験しても，熟達度のレベルについて，個別の試験の独自性や特質を勘案した上で，同じ指標に基づいた教育的なフィードバックを得ることができる。

（２）「マニュアル（Council of Europe 2009，2011）」における対応付けの手続き

Council of Europe (2009，2011)では，CEFR の尺度への対応付けのための次の５つの手続きを示している。

①　Familiarisation（習熟化：CEFR への理解を深める）

対象となる資格・検定試験の対応付けを行う専門家集団（パネル・メンバー）に対し，CEFR，そのレベル区分，言語能力記述文への理解を深める研修を行うこと。

対象となる試験の対応付けを行う委員に対し，CEFR レベルを理解するための
トレーニングを実施する。トレーニングは，事前課題とワークショップに分けられる。トレーニングには以下の a）から i）がある。

事前課題：

a）CEFR の第 3 章第 6 節を読んで各レベルの弁別的特徴を理解する。この際，各レベルの言語活動だけでなく，機能，概念，文法，語彙などの例示尺度についても十分理解する。

b）コーディネーターによって作成された，対象となる試験を CEFR と対応付けする上で必要となる観点（第 3 章～第 5 章の各節末の問いより抜粋）をまとめたチェックリストを確認する。

c）CEFTrain（http://www.helsinki.fi/project/ceftrain/）にアクセスし，各レベルの弁別的特徴を示したパフォーマンスに実際に触れ，言語能力記述文の分析を通して CEFR のレベルをさらに深く理解する。

ワークショップ（約 3 時間）：

＜導入活動＞

d）Council of Europe（2009）付録 A 1 の表（CEFR 第 3 章第 6 節の短縮版）を用いて，レベルの異なる言語能力記述文をレベル順に並べ替える活動を行う。

e）CEFR（2001）に収録されている「自己評価表」に基づいて自分ができる外国語について自己評価を行う。さらに，外国語能力の質的側面に関して，「話し言葉の評価表」または「話し言葉の流ちょうさ」や「文法的正確さ」についての言語能力記述文を用いて自己評価を行うこともできる。その後，他の参加者との共有・議論を行う。

＜言語能力記述文の質的分析＞

f）CEFR（2001）に収録されている言語能力記述文を，レベルを伏せてバラバラにした上で並び替えてレベル付けを行い，なぜそのレベルをつけたのかについてグループで検討を行う。（いくつかのカテゴリーにまたがる場合は，言語能力記述文の合計が 40 を超えない程度とする）

g）CEFR（2001）に収録されている「自己評価表」に用いられている個々の言語能力記述文をバラバラに切り離しレベルを伏せた切片を準備し，それらを正しい位置に配置しなおす活動を行う。

＜産出技能評価の準備＞

h）CEFR の言語能力評価基準表の穴埋めまたは並べ替えタスクを行う。
CEFR の「話すこと」から始めるのであれば，Council of Europe（2009）の付録 C 2 の表，「書くこと」から始めるのであれば，C 4 の表を用いる。（対応付けの対象となる試験に産出技能の評価がない場合でも必ず行うこと）

i）ビデオに撮られた学習者のパフォーマンスを用いて，CEFR レベルを説明する。

② Specification（明確化：対象となる資格・検定試験の自己点検）

資格・検定試験の問題内容や問題タイプについての自己点検を行い，当該試験の出題範囲およびレベルが CEFR と対応付けられること。また，CEFR と対応付かない領域について記述をすること。さらには，内容分析に基づき，CEFR の言語能力記述尺度を用いた当該の試験のプロフィールを描くこと。

自己点検にあたっては，Council of Europe（2009）付録のセクションＡ２の書式のチェックリストを利用して，対応付けの対象となる試験の内容分析を行う。セクションＡ２の書式は全部で 24 あり，内容は以下の通りである。

A1-7：対応付けの対象となる試験の概要に関する書式

A8：対応付けの対象となる試験の CEFR レベルの最初の推定に関する書式

A9-22：対応付けの対象となる試験問題内容に関する書式（A9-18：言語コミュニケーション活動および A19-22：言語コミュニケーション能力）

A23：対応付けの対象となる試験の CEFR と対応付けられた出題範囲とレベルの主張のためのプロフィールの図示に関する書式（必ずしも試験内容の下位分類名と一致している必要はない）

A23：対応付けの対象となる試験の CEFR レベルの最終的な推定に関する書式（Ａ８の書式と異なる推定になった場合はその理由についても明記する）

③ Standardisation training and benchmarking（標準化トレーニング/レベル設定）

パネル・メンバーが基準設定（資格・検定試験のスコアを CEFR に対応付けること）を行うため，試験課題と実際のパフォーマンス例に基づいて，パネル・メンバーの間で CEFR レベルに関する一貫した共通認識を得ること。

話すこと，書くことの実際のパフォーマンスについて，Council of Europe（2009）付録のセクション C1～C4 の評価表を用い，次の三つの段階に分けてレベル判定のトレーニングを行う。その後，対応付けの対象となる試験に関するレベル付けがされていないパフォーマンスの判定を行い，合意形成を行う。また，評価者間や評価者内の信頼性の分析も行う。

第一段階：レベルが確定しているパフォーマンスについての解説を C2，C3（話すこと）および C4（書くこと）の評価表を用いてコーディネーターが行う。C1 の全体尺度を最初に用いてもよい。

第二段階：レベルが確定しているパフォーマンスの判定をコーディネーターのアドバイスを受けながらグループで行う。

第三段階：レベルが確定しているパフォーマンスの判定を個々に行う。

リスニング，リーディング，言語能力についても同様の段階を踏んでトレーニングを行う。出題されるテキストのレベルそのものではなく，問題の難易度との組み合わせにより，受検者の能力を位置づけることに注意する。

④　Standard setting Procedures（基準設定手順）

パネル・メンバーがグループでの数次の審議を経て資格・検定試験のスコアを CEFR の段階別表示に位置付けること。

いくつかの統計的な手法を用いて，受験者のデータを CEFR のレベルに分割し，対応付けの対象となる試験におけるスコアのそれぞれのレベルの境界を明らかにする。この作業に使用する統計の手法は以下の通りである。

1）タッカー・アンゴフ法
2）Yes-No 法
3）応用タッカー・アンゴフ法
4）対照的なグループ法
5）境界線上のグループ法
6）Body of Work 法
7）アイテム-言語能力記述文照合法
8）バスケット法
9）ブックマーク法
10）ブックマーク法 Cito 変法

⑤ Validation（検証）

上記①～⑤の手続きが適切に行われているか，質的，量的な方法に則り継続的に検証する。

（３）国内の外国語試験と CEFR の尺度との対応付けの事例

① 公益財団法人日本英語検定協会

Dunlea, J. (2009,2010)，公益財団法人日本英語検定協会（2018）では，欧州評議会が示している CEFR の尺度への対応付けの手法を用いて英検の各級及びライティングスコアと CEFR レベルとの対応付けを示している。

② 一般財団法人進学基準研究機構
(Center for Entrance Examination Standardization(CEES))

英語コミュニケーションテスト GTEC においては，2016 年度から 2017 年度にかけて CEFR の尺度との対応付けを行い，Pre A1/A1，A1/A2，A2/B1，B1/B2，B2/C1 の各閾値を設定した。2019 年度以降，GTEC 受検者や教師へのフィードバックとして CEFR を用いることを決定している。

６．　社会で活用される日本語能力の判定試験に求められる要素[6]

（１）試験開発に関する基本的な考え方

試験は目的に応じて，開発すべきものであり，本報告は大規模試験において最低限必要となる要素を示すものである。それぞれの概念の説明の下の項目はチェックリストである。

① 有用性（usefulness）

テストは妥当性，真正性，信頼性，波及効果などのさまざまな観点から評価されるが，そのテストの総合的な価値を個々の観点から見た価値の総和として捉える概念である。

② 妥当性（validity）

テストが測定目的とした能力や特性（一般化して「構成概念」と呼ばれる）を確かに測定しているか否かを表わす概念で，「構成概念妥当性」を中心に据えて，それを確認する方法により妥当性の異なる側面が強調される。

＜試験作成・評価に関する妥当性＞

- □ 測ろうとしている言語能力を測っているか。
- □ 試験作成の過程で測るべき知識や能力の一覧を作成しているか。
- □ 測ろうとしてる言語能力を適切に評価できるような採点を行っているか。
- □ 採点基準に基づいた適切な証拠が得られるような問題となっているか。
- □ 測ろうとしている言語能力に対して，設問数のバランスの偏りがないか。また，試験全体の構成概念のバランスと合致しているか。
- □ 測ろうとしている言語能力と直接関連しない設問や内容が含まれていないか。

＜試験実施に関する妥当性＞

- □ 受験者の身体的，心理的，経験的な特徴に配慮した試験になっているか。
- □ 試験内容が受験者間の公平性を保つものになっているか。
- □ 試験内容や能力基準に関する十分な情報が受検者に与えられているか。
- □ 試験の指示や説明が問題の意図を十分に伝えるものとなっているか。
- □ 試験解答のプロセスが測定意図と照らし合わせて不適切なものとなっていないか。
- □ 特別な支援を必要とする受験者に対する配慮がなされているか。
- □ 試験内容についての情報，受験者情報の管理が適切に行われているか。

---

[6] 各要素の説明については野口・大隅（2014）pp.11-23 の説明をもとに一部修正

③　真正性（authenticity）

試験の課題項目がその試験で測定しようとしている目標言語使用領域における現実の課題をどの程度反映しているかの度合いをいう。例えば，読解力を測定する場合には，現実社会で実際に使用された文書を問題文とすることが望ましい。ただし，外国語学習の初級レベルでは既習の語彙・文法・漢字・言いまわしなどに配慮した文章を新たに書き起こすこともある。

□ 設問はどれくらい現実の言語使用場面を反映しているか。
□ 一定のレベル以上の試験においては，現実社会で実際に使用されている文書を活用できているか。

④　信頼性（reliability）

テストの測定精度を表わす概念で，そのテストの測定結果，すなわち，得点に含まれる測定誤差が小さいほど，そのテストの信頼性が高いという。一般に，測定の標準誤差や信頼性係数で表わされる。

□ 何度実施しても同じ結果を得ることができるか。
□ 適切な統計手法を用いて「内的一貫性」，「項目弁別力」等についての検証を行っているか。
□ 対面式の会話試験などで受験者のパフォーマンス能力を測る場合，試験官の間での評価のばらつきがないか，また，試験官のパーソナリティや印象が受験者のパフォーマンスに影響を与えていないか。
□ パフォーマンス能力の測定において，試験者内での評価のばらつきがないかの検証を行っているか。
□ 試験問題の項目困難度のバランスが想定される受験者の能力帯と合っているか。

⑤　実行可能性（feasibility）

テスト開発及びその後の安定的かつ継続的運営を可能にするに関わる人的・経済的資源に関連し，テストを物理的・経済的に成り立たせるための前提条件のことをいう。具体的には，問題開発，テスト実施（採点，結果の通知），データの分析などテストの実施及び結果の検証などに関する一連の流れを継続的に実行可能かということを問題にする。

□ 設問作成にどれくらいの時間がかかるか。
□ 実施にどれくらいの手間で実施できるか。

□ 採点にはどれくらいの時間がかかるか。
□ 上記を勘案して，安定的に実施できるのは年間にどれくらいの頻度になるか。
□ 受験者の受験環境が過度に負担をかけるものとなっていないか。

⑥ 波及効果（washback effect）

試験の内容が受験者や教師，教育機関，企業，それら関係者を含む社会に与える影響のことを言う。例えば，外国語教育機関が外国語試験の出題傾向に合わせて学習内容やカリキュラムを決めるなど。

□ 教育機関のカリキュラム改善に役に立っているか。
□ 受験者に学習方法の改善を促すフィードバックを与えることができているか。
□ 共通参照枠に照らし合わせるなど，透明性の高いフィードバックを与えているか。
□ 試験結果を解釈する十分な情報が与えられ，現実のコミュニケーション場面での言語使用との対応を示しているか。
□ 全体的なレベルだけでない，個々の言語コミュニケーション活動や言語コミュニケーション能力に関する学習者のプロフィールを示しているか。
□ 初級レベル等での学習上の配慮など，真正性から逸脱する点について，受験者に情報が与えられているか。

（２）社会的ニーズに応える日本語の能力判定の在り方について

○ 日本語による言語活動のうち，特に「話すこと（やりとり）」「話すこと（発表）」「書くこと」）の言語能力を測定するテストの開発が求められている。さらには，書くことにおけるオンラインも含めたやり取りの言語能力の判定の必要性も高まっている。

○ 幅広い受験ニーズに応えるため，従来型の紙ベースの試験のほかに，CBT（Computer-Based Testing）による試験の実施が求められている。

○ 「日本語教育の参照枠」に基づき，今後，日本語能力の判定が必要となる外国人材の活動分野や業種等による分野別の日本語能力の評価が行われるようになると想定されるが，試験実施機関側が試験により判定する日本語の分野やレベルを社会に広く明示するようになると良い。

○ 「話すこと」「聞くこと」などの言語活動別に求められる能力レベルが示されることにより，日本語学習及び日本語の能力判定に有効に活用されることが望まれる。

○ 社会的なニーズに応えられる日本語能力の判定試験を開発し，安定的に実施していくためには，試験開発に関する専門人材の育成が不可欠である。

参考文献

○国際交流基金（2016）「JF 日本語教育スタンダードに準拠ロールプレイテストテスター用マニュアル」https://jfstandard.jp/roleplay/ja/render.do （令和２年８月 25 日閲覧）

○日本英語検定協会（2018）「英検ライティングスコアと CEFR レベル対応付け調査報告書」https://www.eiken.or.jp/eiken/group/result/pdf/eiken-score-cefr.pdf （令和２年８月 25 日閲覧）」

○野口裕之・大隅敦子（2014）『テスティングの基礎理論』研究社

○文化庁国語課（2012）「「生活者としての外国人」に対する日本語教育における日本語能力評価について」
https://www.bunka.go.jp/seisaku/kokugo_nihongo/kyoiku/nihongo_curriculum/index_4.html
（令和２年８月 25 日閲覧）

○Center for Entrance Examination Standardization(2018)「GTEC スコアと CEFR J レベル関連付け調査報告」
https://www.benesse.co.jp/gtec/schoolofficials/research/pdf/doc-2018-01.pdf
（令和２年８月 25 日閲覧）

○Council of Europe (2001) Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge University Press. （吉島茂・大橋理枝 訳・編（2014）「外国語の教育Ⅱ　外国語の学習，教授，評価のためのヨーロッパ共通参照枠（追補版）」朝日出版社）https://www.goethe.de/ins/jp/ja/spr/unt/kum/ger.html （令和２年８月 25 日閲覧）

○Council of Europe (2009) Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR)
http://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680667a2d（令和２年８月 25 日閲覧）

○Council of Europe (2011) Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR) Highlights from the Manual.
https://www.ecml.at/Portals/1/documents/ECMLresources/2011_10_10_relex._E_web.pdf?ver=2018-03-21-100940-823（令和２年８月 25 日閲覧）

○Council of Europe (2018) CEFR Companion Volume with New Descriptors
https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989
（令和２年８月 25 日閲覧）

○Dunlea, J. (2009,2010)「英検と CEFR との関連性について研究プロジェクト報告」
https://www.eiken.or.jp/center_for_research/pdf/market/report_02.pdf（令和２年８月 25 日閲覧）

以　上