ライティング評価における評定者の行動分析と評価基準の妥当性検証

－思考表出法による評価行動の分析－

山梨県立大学　杉田由仁

研究の経緯

- 第二言語知識体系：規則に基づく体系と記憶に基づく体系(二層性の体系)、トレードオフの関係(Skehan 2001)
- 「書く」領域における言語運用能力の構成概念：規則に基づく体系への依存度が高い“accuracy,”記憶に基づく体系への依存度が高い“communicability”

テスト開発と調査

- 評価タスク・評定尺度の開発と調査(2008年1月～2010年8月)
- ２種類の評価タスク
  「自己紹介の手紙(100～120語)」20分
  「ディスカッションのためのメモ」10分
  辞書の使用は認めない

ライティング・タスク　1

- You are going to stay with Parker Family in Britain this summer. Write a 100-120 word letter introducing yourself to your host family. Before writing, think of the following topics.
- Your name and age
- Your job, major in school

ライティング・タスク　1

- Your family and pets
- Your interests and hobbies
- Your favorite places, foods, activities
- Your experience in traveling abroad
- Some things you want to do while you are in Britain

今回の研究対象となる評価基準

タスク１の評価基準

評価対象となる言語能力特性：Accuracy　(grammar, organization, vocabulary, rhetoric など、言語の形式的側面における運用能力の正確さ)	
文章構成力 (Organizational skills)：読み手に内容を正確に把握させるために論理的に文章を組み立てる力	**言語的正確さ** (Linguistic accuracy)：語彙や文法、スペル、句読法などにおける誤り
・文章の構成および展開がうまくできている ・論理展開の方法が適切で説得力がある ・さまざまな連結詞の使用により、文章構成が明確である	・部分的に誤りはあるが、語彙使用が適切である ・主語と動詞の一致、時制、単数・複数、語順および語法、冠詞、代名詞、前置詞の使用にほとんど誤りがない ・スペル、句読法、大文字使用、段落分けの仕方にほとんど誤りがない

ライティング・タスク 2

- You will have 10 minutes to make notes about the following discussion topic, "Why do you study English?" In order to prepare for the discussion, think of answers to the question as many as possible and write them as "To travel abroad."

今回の研究対象となる評価基準

タスク２の評価基準

評価対象となる言語能力特性：Communicability (意味・内容の伝達を重視し、言語による効率的な情報伝達を行うことのできる能力)	
伝達内容の質 (Communicative quality)：書かれている内容を、読み手が支障なく、明瞭に理解することができる	**情報伝達の効果** (Communicative effect)：与えられた課題に対して適切かつ十分に、関連性が明確である考えが効果的に示されている
・言語使用能力が確かなものであることがわかる ・自分の考えを表現したり、意図を伝えることのできるすぐれた構文力・語彙力がある	・与えられた課題に対してそつ無く回答している ・課題に対する関連性が十分にある考えが数多く提示され、効果的に内容が伝えられている

研究目的

- 思考表出法により、評定者の評価プロセスにおける行動基準の分析を行う。
- 評価プロセスにおける評定者の行動基準の分析に基づき、開発されたライティングテストにおける評価基準の妥当性検証を行う。

研究方法：分析データの収集

- 第１次データ収集(2008年8月)
  対象：山梨県の公立中学校において英語科を担当する5名の現職教師
  依頼した評定作業の内容：
  タスク1 (accuracy) の評定
  タスク2 (communicability) の評定
  全体的印象による総合評価

研究方法：分析データの収集

- 第2次データ収集(2010年8月)
  対象：第１次データ収集(Session 1)に参加した5名の中学校教師の内1名
  依頼した評定作業の内容：
  Session 1と同一の評価対象データを思考表出法により評定する

分析方法

- FACETS 分析により、Session 1および2における評定結果の信頼性を検証する
- 思考表出法によって得られたプロトコルデータに基づき、評定者の評価プロセスにおける行動基準の分析を行う
- 両セッション間の評定の変化に着目し評価基準の妥当性を検討する

結果：FACETS 分析

- 被験者の能力

  20名の被験者が、−14〜14 (logits)までの28段階に評定された。

  2回の評定セッションにおいて評定が一致した割合(inter−rater agreement)は64.9%であり、“independent experts”として評定が行われた(Linacre, 2007)

結果：タスクの難易度

タスク	1	2	印象
難易度	0.52	−0.35	−0.17
誤差	0.41	0.44	0.43
Infit	0.92	0.90	1.00

→ロジット得点による難易度：$\chi^2(2)=2.4$, p=.29
タスクの評定値および全体的評価に、難易度の違いによる影響は認められない 。Infit数値も平均(0.94)より標準偏差(0.04)の2倍値の範囲内であり、一貫した評定が行われた。

結果：評定セッション

セッション	測定値	期待値	差	バイアス	Zスコア
S1 (A)	59	59.5	−.03	−.17	−.29
S1 (C)	60	62.0	−.10	−.71	−1.21
S1 (Ｉ)	60	61.5	−.08	−.52	−.90
S2 (A)	60	59.5	.03	.17	.29
S2 (C)	64	62.0	.11	.88	1.22
S2 (Ｉ)	63	61.5	.08	.59	.90

→Zスコアは「有意でないバイアス(-2<Z<2)」の範囲内にあり、評定者は独自のバイアスを持たずに評定できた。

結果：プロトコルデータの分析

評定基準 (評価の観点) 出現頻度：Accuracy

コード	カテゴリー	サブ・カテゴリー	評定基準 (評価の観点)	頻度(%)
A11	Accuracy	文章構成力	文章の構成・展開	18 (25.0)
A12	Accuracy	文章構成力	論理・展開方法の適切さ	11 (15.3)
A13	Accuracy	文章構成力	連結詞の使用	2 (2.8)
A14	Accuracy	文章構成力	その他(語数)	10 (13.9)
小計				41 (56.9)
A21	Accuracy	言語的正確さ	語彙使用の適切さ	7 (9.7)
A22	Accuracy	言語的正確さ	文法的誤り	12 (16.7)
A23	Accuracy	言語的正確さ	表記上の誤り	3 (4.2)
A24	Accuracy	言語的正確さ	その他	9 (12.5)
小計				31 (43.1)
合計				72 (100)

結果：プロトコルデータの分析

評定基準 (評価の観点) 出現頻度：Communicability

コード	カテゴリー	サブ・カテゴリー	評定基準 (評価の観点)	頻度(%)
C11	Communicability	伝達内容の質	言語使用能力	2 (3.3)
C12	Communicability	伝達内容の質	考えや意図の明瞭さ	18 (29.5)
C13	Communicability	伝達内容の質	構文力・語彙力	11 (18.0)
C14	Communicability	伝達内容の質	その他	2 (3.3)
小計				33 (54.1)
C21	Communicability	情報伝達の効果	そつのなさ	13 (21.3)
C22	Communicability	情報伝達の効果	関連性	3 (4.9)
C23	Communicability	情報伝達の効果	情報量(理由の数)	10 (16.4)
C24	Communicability	情報伝達の効果	その他	2 (3.3)
小計				28 (45.9)
合計				61 (100)

結果：評定傾向の変化(Communicability)

評定傾向の変化に影響を与えたカテゴリー

評定	Session 1 → Session 2	評定傾向の変化に影響を与えたカテゴリー
S2	3 → 4	C23
S12	4 → 5	C12, C23
S14	3 → 4	C12, C23
S20	2 → 3	C11, C12, C21

→両セッションにおける評定傾向の変化に影響を与えたカテゴリーは、C12(考え方や意図)およびC23(情報量)であり、C12・C23には評定における「葛藤」が派生

結果：評定結果の変化(Accuracy)

評定結果の変化に影響を与えたカテゴリー

評定	Session 1 → Session 2	評定結果の変化に影響を与えたカテゴリー
S10	3 → 4	A12, A21, A22
S12	4 → 5	A11, A12, A21, A22, A24
S13	3 → 4	A11, A12, A14
S18	4 → 5	A11, A12, A24
S1	4 → 3	A11, A12
S2	4 → 3	A22, A24
S17	3 → 2	A11, A12, A22
S20	4 → 3	A11, A24

→上昇：A12(論理・展開方法の適切さ)
下降：A12（論理・展開方法の適切さ）A22(文法的誤り)
A12に対してA14(その他：語数)による葛藤の調停

考察：評定者の役割（Lumley, 2002）

- 評定者は「文章の構成・展開」「文法的誤り」「考え方や意図の明瞭さ」「情報量(理由の数)」に着目し、バランスの取れた評価行動を行っていた
- 評価プロセスにおいて派生した葛藤に対して一貫性のある調停を行うことができた

→ 妥当な評価基準

考察とまとめ

- FACETS分析による信頼性検証において「評価対象となる言語能力特性」に基づいて評定を行っていたことが裏付けられた
- Session 1 における評定経験が評定トレーニングとして機能し、評定に関わる自己判定についての説明を尺度に基づいて行うことができた

主要引用文献

Linacre, J.M. (2007). *A user's guide to FACETS*.

Lumley, T (2002). Assessment criteria in a large-scale writing test. Language Testing 19, 246-76.

Skehan, P. (2001). *A cognitive approach to language learning*. Oxford University Press.