

一般化可能性理論を用いた外国人児童生徒のための JSL 対話型アセスメント DLA 〈話す〉の検討

森 重 里 保・渡 部 倫 子

Utilizing Generalizability Theory in the Evaluation of Dialogic Language Assessment

Riho MORISHIGE, Tomoko WATANABE

キーワード：DLA (Dialogic Language Assessment), 信頼性, 評価, 一般化可能性理論

1. はじめに

DLA (Dialogic Language Assessment, 対話型アセスメント) は、文部科学省の「外国人児童生徒の総合的な学習支援事業」(2010~2012)の一環で、「学校において利用可能な日本語能力測定方法の開発」として開発されたものである。DLA では、テスターがCLD児 (Culturally and Linguistically Diverse Children)¹⁾と1対1で対話をしながら、子どもの言語能力を把握すると同時に、彼らにとってどのような学習支援が必要であるか、検討することを目的としている。このうち、DLA 〈話す〉では、話す力を基礎会話・対話・認知の3面から捉え、場面に依存して対応できる言語能力から、認知力を必要とする言語能力までを含む幅広い力を測ることができる。評価は、日常の授業態度も加味しつつ「JSL 評価参照枠〈話す〉」を参照し、ステージ1 (初期指導段階) からステージ6 (支援付き自律学習段階) までの6段階で行う。しかし、DLA で用いられる評価基準がどれだけ信頼性の高いものであるのか、どのように評価すればより信頼性の高い評価ができるのかという点については未だ十分な議論がされていないのが現状である。そこで本研究では、一般化可能性理論を用いて上記の問題に対する検討を行う。研究課題は次のとおりである。① DLA 〈話す〉の評定に、児童、評定者、評価項目というそれぞれの要因や交互作用がどの程度影響を及ぼすのか。② DLA 〈話す〉の評定において十分な信頼性を得るためにはどの程度の評定者数、項目数が適切か。

2. 先行研究

一般化可能性理論は、一般化可能性研究 (G 研究) と決定研究 (D 研究) から構成される。G 研究から

は各分散成分が推定され、D 研究からは具体的な条件 (項目数や評定者数) における一般化可能性が検討される (池田, 1994)。池田 (1994) によると、一般化可能性理論は、あるテストで1題につき何人くらいの評定者を用意したらよいか、受験者に与える課題を1人何題くらいにしたらよいか等、効果的なテスト計画を立てるのに必要な情報を与えてくれる。

佐々木・村木 (2005) は、大学の授業で実施した口頭発表を対象に、どれだけの項目数と評定者数を用意すれば信頼性のある評価が行えるかについて、一般化可能性理論に基づいて検証している。一般化可能性係数が0.8以上であることを信頼性の基準として設定し、分析を行ったところ、評定者数が少ない場合には項目数を増やしても信頼性のある評価が期待されないことが明らかになった。また、学校現場では評定者を増やすことが簡単でない場合も多いため、単に評定者数を増やすよりも、5名程度の評定者のもとで項目数を増やした方が信頼性の向上に効果的であることが示唆された。このように、一般化可能性理論の結果と学校現場とを併せて考えることで、よりよい評価を行うための情報を得ることができる。

坂野 (2008) は、61人の評定者が6人の中国人話者の口頭テストを評価したデータを使用し、日本語の口頭プレースメントテスト分析における一般化可能性理論の可能性を検討した。その結果、評定者一人、設問数一問というテスト形式では、受験者の能力以外の要因により、テスト得点に誤差の生じる可能性が少なくないことが明らかになり、2名の評定者と2問の設問が必要であると示唆された。

外国人児童生徒数が増加している中で、学校では彼らの日本語能力について信頼性のある評価を行うことが期待されるが、それに関する具体的な提案はいまだ少ないのが現状である。以上より、本研究の

研究課題を以下の2つに設定した。①DLA〈話す〉の評定に、児童、評定者、評価項目というそれぞれの要因や交互作用がどの程度影響を及ぼすのか。(G研究) ②DLA〈話す〉の評定において十分な信頼性を得るためにはどの程度の評定者数、項目数が適切か。(D研究)

3. 分析方法

3.1 分析対象

分析の対象は、A小学校に通う小学3年生15名を対象に行った、DLA〈話す〉のデータを評価者2名が評価したものである。なお、評定者は2名とも、以前にDLAのワークショップを受講したことがある。評価基準はDLA〈話す〉で用いられる、JSL評価参照枠〈話す〉を使用した(表1)。これは、日本語能力のうち、特に話す力の発達段階を6つのステージに分けて、総合的かつ多面的に記述したものである。ステージ6に近づくほど、能力が高いと判断される。

3.2 分析方法

統計計算のための言語・環境 R3.1.2を用い、一般化可能性理論による分析を行った。本研究では、JSL評価参照枠〈話す〉(表1)の6つの項目数(「話の内容・まとめ」「文・段落の質」等)を1から15と変動させたときの一般化可能性係数のほか、評定者数を1名から7名としたときの一般化可能性係数についても調べた。分析の際は2相完全クロス計画(対象児童×項目×評定者)を用い、2つの相(項目および評定者)のどちらも変量モデル²⁾とした。なお、本分析では、佐々木・村木(2005)と同様、信頼性の基準として、一般化可能性係数が0.8以上であることを条件とした。

4. 分析結果

4.1 G研究の結果

分析の結果得られた、分散成分の推定値とその割合(カッコ内の数字)を表2に示す。この表2の変動要因のうち、「児童×評定者×項目」の分散成分の推定値の割合(52.0%)が全変動要因中の大部分を占めることから、タスクの難易度や評価項目の内容など、各変動要因(児童、評定者、評価項目)では説明できない要因が多く存在すると考えられ

表1 JSL 評価参照枠〈話す〉
(文部科学省, 2014: 36をもとに作成。6~1の数字はステージを示す)

話の内容・まとめ	6	年齢相応の教科内容と関連した認知タスクがこなせる
	5	まとまった話が1人できる
	4	文法的正確度が高い
	3	年齢相応の教科学習語彙が使える
	2	発音が自然で、流暢度が高い
	1	自分から進んで発言し、会話を自らリードできる
文・段落の質	6	まとまった話が1人できる
	5	ある程度まとまった話ができる
	4	文を生成し、ある程度連文ができる
	3	単文レベルの応答ができる
	2	二語文
	1	一語文
文法的正確度	6	文法的正確度が高い
	5	文法的正確度がある程度高い
	4	連文レベルで誤用がほとんど目立たない
	3	単文は生成できるが、助詞や活用などの誤用が目立つ
	2	語順が乱れ、活用が不正確
	1	単語レベル
語彙	6	年齢相応の教科学習語彙が使える
	5	教科学習語彙がある程度使える
	4	日常語彙が使える
	3	身近な日常語彙が使える
	2	基礎語彙が使える
	1	わずかな基礎語彙が使える
発音・流暢度	6	発音が自然で、流暢度が大変高い
	5	発音が自然で、流暢度が高い
	4	発音が自然で、流暢度がある
	3	流暢度が低い
	2	流暢さなし
	1	流暢さなし
話す態度	6	自分から進んで発言し、会話を自らリードできる
	5	様々な会話に積極的に参加することができる
	4	聞かれた質問に答えることができる
	3	聞かれた質問にある程度答えることができる
	2	定型表現や知っている単語でコミュニケーションをとろうとする
	1	ジェスチャーや表情でコミュニケーションをとろうとする

表2 分散成分の推定値と一般化可能性係数

変動要因	分散成分推定値
児童 (s)	0.17 (8.5%)
評定者 (r)	0.11 (5.6%)
項目 (i)	0.10 (5.2%)
児童×評定者 (sr)	0.11 (5.6%)
児童×項目 (si)	0.03 (1.5%)
項目×評定者 (ir)	0.42 (21.4%)
児童×評定者×項目 (sri)	1.02 (52.0%)
一般化可能性係数	0.53

る。また、「項目×評定者」の交互作用の分散成分の推定値の割合(21.4%)から、今回の評定者には評価項目の捉え方に比較的大きな違いがあったことが示された。

4.2 D研究の結果

ここでは、G研究で得られた分散成分の推定値を使用して一般化可能性係数を算出し、評定者数と信頼性の関係のシミュレーションを行った。その結果を図1に示す。

本分析では、信頼性の基準として、一般化可能性係数が0.8以上であることを条件としたが、一般化可能性係数は、評定者が1名から4名の場合、すべて0.8未満であった。評定者が5名の場合、項目数が11のとき0.79、項目数が12のとき0.80となった。また、評定者が6名の場合、項目数が8のとき0.79、項目数が9のとき0.81となった。さらに、評定者が7名の場合、項目数が6のとき0.79、項目数が7のとき0.80となった。

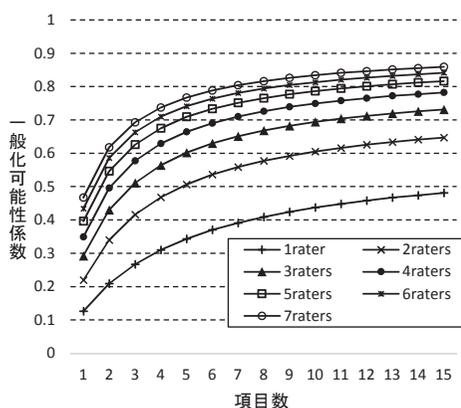


図1 項目数と評定者数を変化させたときの一般化可能性係数

5. 考察

本調査の結果から、小学3年生を対象としたDLA〈話す〉の評価において、信頼性のある評価を行うためには、評定者が5名の場合に12項目以上、評定者が6名の場合に9項目以上、評定者が7名の場合には7項目以上が必要であること、また、評定者数が4名以下の場合、項目数を増やしても信頼性のある評価が行われないことが明らかになった。ここから、佐々木・村木(2005)と同様、評定者数が少ない場合、項目数を増やしても信頼性のある評価は得られにくいことが示唆された。ただし、授業や児童、さまざまな行事などに追われる学校現場では評定者数を増やすことが困難であることも多いため、単に評定者数を増やすのではなく、佐々木・村木(2005)で述べられているように、適当な評定者数のもとで項目数を増やした方がより効果的だと考えられる。本研究の結果から言えば、7名の評定者で7項目を用いる(一般化可能性係数0.80)よりも、5名の評定者で12項目を用いる(一般化可能性係数0.80)方がより現実的である。そのため、現在JSL評価参照枠〈話す〉で設定されている6つの発達段階に新しい記述を加えたり、さらに細分化するなどして評価基準を明確にする必要がある。

4.1でも述べた通り、今回の分析で設定した変動要因では説明することのできない部分が多くあった。この点に関して、本研究で評価対象としたタスクが、対象児童の言語能力を十分に引き出しうる内容であったかどうかについては今後さらに検討する必要がある。さらに、「まとまった話が1人でできる」のひとつ下のステージが「ある程度まとまった話が1人でできる」等、評価基準に曖昧なものが多く、評価に大きな影響を与えていると考えられる。また、今回は対象児童の日本語習熟度レベルについて考慮できていなかったため、今後は児童間の習熟度についても考慮し、分析を行わなければならない。本研究では評価参照枠〈話す〉について分析を行ったが、今後は評価参照枠だけでなく、診断シート³⁾や質的評価シート⁴⁾についても検討することで、評価基準を改善していく必要がある。

謝辞

本研究は、2016年度JSPS科研費25284096(基盤研究(B))『アーティキュレーションを保証する言

語能力アセスメント実施支援システムの構築』研究
代表者：渡部倫子)の助成を受けたものである。

注

- 1) 多種多様な文化的・言語的背景をもつ子どものことを指す(カミンズ・中島, 2011) 32。
- 2) 池田(1994)によると, 変量モデルでは, 取り上げる課題や評定者が, より大きな許容観測母域からのサンプルであるとみなす。つまり, 考えられる多くの処理水準の中からランダムに水準を選んだと考える。
- 3) 各技能の測定結果を記入するためのもので, 〈話す〉では, 正答・無回答のどちらかで表される。
- 4) DLA 〈話す〉での子どもの発話を5(とてもよい)・3(ふつう)・1(もう少し)の3段階で表し, 質的に評価するためのものである。テストターが評価に揺れる場合は, 中間を選択し, 4点または2点をつけてもよいとされている。

引用文献

- 池田央(1994)『現代テスト理論』朝倉書店
- 佐々木典彰・村木英治(2005)「口頭発表の評価における信頼性—一般化可能性理論を用いて—」『教育情報学研究』3, 1-4
- ジム・カミンズ・中島和子(2011)『言語マイノリティを支える教育』慶應義塾大学出版会
- 坂野永理(2008)「一般化可能性理論による日本語口頭ブレースメントテストの検討」『日本テスト学会誌』4(1), 23-32
- 文部科学省初等中等教育局国際教育課(2014)「外国人児童のためのJSLアセスメントDLA」
(http://www.mext.go.jp/a_menu/shotou/clarinet/003/1345413.htm)よりダウンロード可能