

日本語版 Can-do Scale はどれくらい正確に日本語能力を測定しうるか
**How accurately does a Japanese version of a Can-do Scale
measure Japanese language ability?**

玉岡賀津雄 (TAMAOKA, Katsuo)

松下達彦 (MATSUSHITA, Tatsuhiko)

元田 静 (MOTODA, Shizuka)

日本語の要旨: 日本語版 Can-do Scale が、どれくらい正確に外国人日本語学習者の日本語能力を測定しうるかを検討した。その結果、「聞く」「話す」「読む」「書く」の4技能からなる20項目の質問すべてについて、項目間に有意な相関が得られた。また、クロンバックのアルファ係数も極めて高い数値であった。しかし、因子分析の結果、「聞く」「話す」「書く」の群と「読む」の群の2つの因子にしか分けられないことがわかった。また、日本語能力テストの得点と日本語版 Can-do Scale で測定した4技能の間には、有意な相関はみられなかった。さらに、4技能から日本語能力を予測する重回帰分析を行ったが、どの技能も日本語能力を予測する有意な変数ではなかった。つまり、自己評価型の質問紙は、日本語の能力を予測できないという結果であった。これは個人の社会・文化的背景、言語不安や性格要因などから、自己の言語能力を過大あるいは過小に自己評価してしまうためであると考えられよう。

日本語のキーワード: 日本語版 Can-do Scale, 日本語能力の測定, 自己評価, 質問紙法, 信頼性と妥当性

英語の要旨: The present study examined how accurately a Japanese version of a Can-do Scale measures actual Japanese language ability of international students. Twenty questions related to the four categories of 'listening', 'speaking', 'reading' and 'writing' had significantly high correlations in every possible combination. A Cronback's alpha coefficient indicated a very high consistency of 20 items. However, the factor analysis identified only two factors: a group of 15 questions regarding 'listening', 'speaking' and 'writing', and a group of 5 questions regarding 'reading'. Correlations between the four categories measured by the Japanese version of the Can-do Scale and a total score of a Japanese language ability test were not significant. In addition, multiple regression

analysis predicting the Japanese language ability by the four categories did not reveal any significant predictor. Thus, it is concluded that the Japanese version of the Can-do Scale cannot properly measure actual Japanese ability. International students learning Japanese are likely to overestimate or underestimate their Japanese ability depending on their cultural/social backgrounds, language anxiety and personality factors.

英語のキーワード: Japanese version of Can-do Scale, measuring Japanese language ability, self-assessment, questionnaire methods, reliability and validity

1. 研究の目的

「日本語のラジオのニュースがわかりますか。」のような Can-do 形式の質問によって日本語能力を測定しようとするのが Can-do Scale である。この種の質問紙には、二つの目的がある (Leblanc & Painchaud, 1985)。それらは、言語能力試験を行う費用の削減と時間の効率性である。仮に、学習者が正確に言語能力を自己判断できるのであれば、Can-do Scale は、能力別クラス配置や進路指導などにも役立つであろう。日本語教育学会 (1999) は、日本語版の Can-do Scale を開発してきた。しかし、報告によると日本語能力と Can-do 形式の質問の回答との相関は必ずしも高くないようである。

日本語版 Can-do Scale のような自己評価型の質問紙の良さを発揮させるには、回答時間を 5 分から 10 分以内におさめなくてはならない。一つの質問に対して 30 秒を要するとすれば、20 問で 10 分となる。つまり、多くとも 20 問くらいの質問数で測定の目的が達せられるような質問紙を作成しなくてはならないのである。これを超えて 10 分以上の時間を要するようであれば、むしろ日本語能力テストを実施する方が効果的であろう。そこで、本研究では、日本語版 Can-do Scale の質問数を 20 問に限定した。その上で、このような時間効率のよい質問紙が、どのくらい正確に外国人日本語学習者の日本語能力を測定するか。つまり、日本語版 Can-do Scale の信頼性 (reliability) と妥当性 (validity) を検討した。

2. 調査 1

信頼性とは、質問項目が一貫しており、また安定して同じように測定していることの間を示している (ブラウン, 1999; 日本語教育学会, 1991)。日本語版 Can-do Scale については、質問項目が一貫して同じように測定していることを検討するために、質問項目全

体について、ピアソンの相関係数、クロンバックのアルファ係数を算出した。また、これは妥当性に属する分析であるが、20の質問項目の因子構造を因子分析で検討した。さらに、「聞く」「話す」「読む」「書く」に関連した言語技能について、その難易度を分散分析で検討した。

2.1 被調査者

日本の大学・大学院で学んでいる193名の留学生(平均年齢が29歳と2カ月、標準偏差が6カ月;女性96名、男性97名)に日本語版 Can-do Scale(英語の翻訳を併記した)を実施した。これらの留学生の日本語能力は、初級レベルから超上級レベルまでで、専門分野は多岐にわたっている。出身国は、中国が82名、韓国が26名、インドネシアが9名、台湾が8名、マレーシアが4名、その他64名である。また、学籍は、学部生が12名、大学院生が110名、研究生が51名、その他が20名であった。

2.2 日本語版 Can-do Scale の質問項目

質問項目の選択には、まず MacIntyre, Noels & Clément(1997)の改訂版の質問項目を翻訳した。しかし、カナダの文化的環境で作られた質問であるため、日本の状況に当てはまらない質問がかなりの数あることが分かった。そこで、単純な翻訳では使えないと判断して、元田(2000)が用いた Can-do Scale を参照した。さらに、日本語教育学会(1999)の質問項目を参照したが、日本語能力試験との相関が低いと判断したため、質問で聞いている日本語能力の内容が簡単すぎて、天井効果が出てしまったのではないかと思われた。そこで、平均値の高い順(易しい順)に並べ、50項目以上の質問から、日本の大学・大学院で通常の授業を受講している学生を想定して、ある程度難易度が高く、日本での生活環境や大学での学習環境に合うと思われる質問項目を日本語版 Can-do Scale として選んだ。その際、「聞く」「話す」「読む」「書く」に関連した4つの技能について5つずつ、合計20項目とした(質問項目の詳細は表1を参照)。そして、それぞれの質問について、5(非常によくできる)、4(よくできる)、3(ある程度できる)、2(あまりできない)、1(ぜんぜんできない)の5段階評価でたずねた。質問項目の平均と標準偏差は表1に報告したとおりである。

表 1 日本語版Can-do Scaleの平均、標準偏差および斜交プロマックス回転後の因子パターン行列および因子間相関

質問項目の内容	標準偏差		第1因子 (F1)	第2因子 (F2)
	平均			
知らない人から電話がかかってきても、何を言っているかすぐにはわかりませんか。	3.259	1.116	.780	.028
テレビのドラマがわかりますか。	3.306	1.116	.782	.109
電車やバスの中で、まわりの日本人が話している内容がわかりますか。	3.212	1.100	.716	.071
く授業やゼミで話されている内容がわかりますか。	3.280	1.097	.655	.177
ラジオのニュースがわかりますか。	2.938	1.202	.536	.387
面接で、質問に適切に答えることができますか。	3.016	1.092	.995	-.122
話 自分の観た映画について、ストーリーをくわしく説明することができますか。	2.860	1.135	.821	.107
す パーティーや公式の席で、スピーチをすることができませんか。	2.845	1.135	.810	-.016
授業やゼミの発表で、自分の考えをじゅうぶんに説明することができますか。	2.922	1.127	.797	.126
授業やゼミなどで、日本人と討論することができませんか。	2.845	1.202	.576	.381
お世話になった人にお礼の手紙を書くことができますか。	3.155	1.083	.747	.061
書 だれかに頼まれて、自分の国の事情などについて書くことができますか。	3.202	1.092	.645	.225
論文などの要約を書くことができますか。	2.870	1.224	.570	.356
く 授業やゼミなどで、ノートをとることができますか。	3.088	1.194	.535	.268
調査や研究などの報告書を書くことができますか。	2.860	1.232	.533	.388
小説を読んで、内容がわかりますか。	2.715	1.249	-.047	.971
読 新聞を読んで、内容がわかりますか。	2.953	1.308	-.021	.950
アパートなどの契約書を読んで、内容がわかりますか。	3.145	1.331	.016	.908
む 研究に必要な本や論文を読んで、内容がわかりますか。	3.052	1.274	.104	.819
電気製品(例えば、CDプレーヤーや電子レンジなど)の使い方の説明書を読んで、内容がわかりますか。	3.301	1.165	.180	.693
因子間相関			F1	F2
				.802

注: n=193.

3. 調査1の分析結果

3.1 質問紙の信頼性

ピアソンの相関係数を算出した結果、20の質問項目間の180種類の組み合わせすべてについて、 $r=.50$ 以上の有意な相関(すべての相関係数が $p<.01$)が得られた。すべての質問項目がお互いに強い関係を持つことが分かった。また、クロンバックのアルファ係数は、「聞く」技能についての質問項目が $\alpha=.94$ 、「話す」が $\alpha=.95$ 、「読む」が $\alpha=.96$ 、「書く」が $\alpha=.94$ で、全体では $\alpha=.98$ という極めて高い数値であった。これは、20項目の質問には内的な一貫性があり、信頼性が高いことを示している。

3.2 質問紙の因子構造

20の質問項目の因子構造を調べるために、因子分析(最尤法による因子抽出法で、Kaiserの正規化を伴うプロマックス法による斜交回転)を行った。その結果、表1に示したように、わずか2つの因子群しか抽出されなかった。第1因子は、「聞く」「話す」「書く」についての15項目であり、第2因子は「読む」の5項目であった。つまり、「読む」ことが他の言語活動と異なる以外は、音声言語に関する「聞く」「話す」と文字言語に関する「書く」ことが一つの因子として見いだされた。これは、「読む」ことが異なる能力として意識されている以外は、「聞く」「話す」「書く」がほぼ類似した能力として測定されていることを意味している。さらに、第1と第2因子間の因子相関が $r=.802$ と高かった。

3.3 言語技能に関する分析

本研究で使用した日本語版 Can-do Scale は、4つの言語技能から構成されている。さらにこれらの技能は、情報の媒介という観点から、「聞く」と「話す」を併せて「音声言語」、「読む」と「書く」を併せて「文字言語」に分けられる。さらに、情報の入力という意味での知覚(perception)に関連する「聞く」と「読む」を併せて「知覚言語」、情報の出力という意味での産出(production)に関連する「話す」と「書く」を併せて「産出言語」という2種類の言語技能に分けられる。4技能をこれらの分類で計算した平均および標準偏差は表2に示した。これらの技能分類から、2(音声・文字)×2(知覚・産出)の分散分析を行った。その結果、知覚言語($M=31.16$, $SD=10.50$)か産出言語($M=29.65$, $SD=10.50$)かの変数に有意な主効果がみられた [$F(1, 192)=19.16$, $p<.001$]。さらに、産出言語のなかの比較も行ったが、「話す」技能($M=14.49$, $SD=5.19$)の方が、「書く」技能($M=15.18$, $SD=5.27$)よりも難しいと感じているという結果であった [$F(1, 192)=11.72$,

$p < .001$ 】。一方、音声言語 ($M=30.48$, $SD=9.84$) か文字言語 ($M=30.34$, $SD=10.61$) には有意な主効果はみられなかった [$F(1, 192)=0.149$, $p=.700$]。両主効果の交互作用は有意であった [$F(1, 192)=27.06$, $p < .001$]。以上の分析結果をまとめると、本研究の被調査者である留学生は、知覚言語よりも産出言語を難しいと、またさらに「書く」技能よりも「話す」技能の方が難しいと感じていた。ただし、言語の媒介が音声であるか文字であるかについては難易度に違いはなかった。

4. 調査2

妥当性とは、質問項目が意図した内容を正確に測定しているかの指標である(日本語教育学会, 1991; ブラウン, 1999)。日本語版 Can-do Scale と日本語能力テストを同一の被調査者に実施し、ピアソンの相関係数を算出した。さらに、日本語版 Can-do Scale で測定した「聞く」「話す」「読む」「書く」の4つの技能得点が、日本語能力を適切に予測しうるかどうかを重回帰分析で検討した。さらに、調査1と同様に、「聞く」「話す」「読む」「書く」に関連した技能について、その難易度を分散分析で検討した。

4.1 被調査者

日本の大学の学部課程で学んでいる34名(平均年齢が23歳9カ月、標準偏差が4歳0カ月; 女性23名, 男性11名)の留学生が調査2に参加した。調査1の被調査者は、調査2には参加していない。出身国・地域の内訳は、中国が20名で圧倒的に多く、台湾が7名、韓国が3名、アメリカ、オーストラリア、ペルー、ロシアがそれぞれ1名である。

4.2 日本語版 Can-do Scale

調査1と同様に表1の20項目からなる日本語版 Can-do Scale を使用した。

4.3 日本語能力テスト

日本語版 Can-do Scale の妥当性の基準として、文字・語彙・文法力を合わせて22問の58点満点 ($M=37.21$, $SD=9.35$) と読解力が14問の42点満点 ($M=26.74$, $SD=8.07$) からなる総合日本語能力(100点満点, $M=63.94$, $SD=14.87$) テストを、全被調査者の34名に実施した。100点満点のテストに対して平均が63.94点で、標準偏差が14.87点であった。63.94パーセントの被調査者(約23名)が平均から±1の標準偏差である49.07点から78.81点の間に入り、残りの約11名がこの範囲から外れている。分散の広がりから考えて、このテストはかなり弁別力があると判断されよう。また、文字・語彙・文法力と読解力のピアソンの相関も有意である ($r=.454$, $p < .01$)。

5. 調査2の分析結果

5.1 日本語能力と言語技能の相関

表2に示したように、日本語能力テストの文字、語彙、文法を合わせた得点、読解の得点、全体の得点と4技能に関する質問の間のピアソンの相関係数を算出した。その結果、どの組み合わせも有意な相関とはならなかった。とりわけ、日本語能力テストが文字を媒介とした問題であったので、Can-do Scaleの「読み」との相関が高くなると予想していたが、実際には、文字・語彙・文法力との相関係数が $r=.235(n.s.)$ 、読解力との相関係数が $r=.279(n.s.)$ 、合計との相関係数が $r=.299(n.s.)$ に留まった。また、文字言語(「読む」と「書く」を合わせた技能)でも、文字・語彙・文法力との相関係数が $r=.032(n.s.)$ 、読解力との相関係数が $r=.205(n.s.)$ 、合計との相関係数が $r=.132(n.s.)$ であった。このように、被調査者34名についての日本語版 Can-do Scale と日本語能力テストのピアソンの相関係数はすべて有意ではなく、Can-do Scale の自己評価の結果が、実際の日本語能力テストの得点を具現化していないことを示唆した。

5.2 日本語能力を予測する言語技能

相関係数だけでは、2つの変数間の関係を調べたに過ぎない。そこで、4つの言語技能(説明変数)から日本語能力を予測する重回帰分析(強制投入法およびステップワイズ法)を行った。その結果、強制投入法では表3に示したように、読解の得点を予測する技能として「話す」技能が有意な説明変数となった。しかし、この結果もステップワイズ法では、有意に読解を予測する変数とはならなかった。ステップワイズ法では、Can-do Scale で測定したどの技能の自己評価も、文字・語彙・文法力、読解力、合計の3種類の日本語能力得点を予測する有意な変数とはならなかった。これらの結果は、日本語版 Can-do Scale が、日本語能力を予測するものではないことを示している。

5.3 言語技能に関する分析

調査1と同様に、2(音声対文字)×2(知覚対産出)の分散分析を行った。その結果、音声言語($M=32.09$, $SD=4.85$)か文字言語($M=32.28$, $SD=5.58$)かには有意な主効果はなく [$F(1, 33)=0.095$, $p=.760$]、知覚言語($M=34.00$, $SD=5.23$)か産出言語($M=30.37$, $SD=5.50$)かについては、有意な主効果がみられた [$F(1, 33)=23.29$, $p<.001$]。両主効果の交互作用は有意ではなかった [$F(1, 33)=2.57$, $p=.119$]。また調査1とは異なり、「話す」と「書く」技能を比較した結果には有意な違いはなかった。これらを要約すると、調査1の結果

表2 日本語版Can-do Scaleによる自己評定と日本語能力テストとのピアソンの相関係数

言語技能の分類	平均	標準偏差	日本語能力テストの得点	
			文字・語彙・文法	読解
「聞く」技能	17.24	2.54	.013	.012
「読む」技能	16.76	3.60	.235	.279
「話す」技能	14.85	3.06	.040	.315
「書く」技能	15.51	2.76	-.066	.066
音声言語(「聞く」と「話す」)技能	32.09	4.85	.119	.213
文字言語(「読む」と「書く」)技能	32.28	5.58	.032	.205
知覚言語(「聞く」と「読む」)技能	34.00	5.23	.168	.198
産出言語(「話す」と「書く」)技能	30.37	5.50	-.011	.208
4技能の合計	64.37	9.80	.084	.222
				合計
				.015
				.299
				.197
				-.006
				.190
				.132
				.213
				.106
				.173

注: n=34. どれにも有意な相関はない。

表3 日本語能力を予測する日本語版Can-do Scaleの言語技能

言語技能	文字・語彙・文法力		読解力		合計				
	β	t値	β	t値	β	t値			
「聞く」技能	-0.004	-0.020	n. s.	-0.131	-0.648	n. s.	-0.074	-0.355	n. s.
「読む」技能	0.353	1.621	n. s.	0.245	1.212	n. s.	0.355	1.709	n. s.
「話す」技能	0.118	0.399	n. s.	0.631	2.288	*	0.417	1.471	n. s.
「書く」技能	-0.344	-1.126	n. s.	-0.488	-1.724	n. s.	-0.481	-1.652	n. s.
重決定係数	$R^2 = .110$		$R^2 = .234$		$R^2 = .190$				

注：表の重回帰分析は強制投入法。n=34。* p<.05。βは標準偏回帰係数。

と類似しており、「話す」(M=14.85, SD=3.06)と「書く」(M=15.51, SD=2.76)に関連した産出言語を、「聞く」(M=17.24, SD=2.54)と「読む」(M=16.76, SD=3.60)に関連した知覚言語よりも難しいと感じているが、言語の媒介が音声か文字かについては難易度に違いはないと感じていることを示している。

5.4 日本語版 Can-do Scale でみる自己の過大および過小評価

Taylor & Brown (1994)によると、第二言語学習者は、基本的に自己を肯定的に見る傾向があり、第二言語習得について自己の言語能力を非現実的なほどに過大評価するようだと述べている。また、逆に「言語の学習や表現に対する不安」(language anxiety)などから自己の言語能力を低く見積もり、適切な自己評価ができない場合もあるという報告や議論もある(e.g., Ferguson, 1978; Horwitz, Horwitz & Cope, 1986)。そこで、まず、日本語能力テストの総合得点(max=100, M=63.94, SD=14.87)を日本語版 Can-do Scale の総合得点(Max=100, M=64.37, SD=9.80)で予測する単回帰分析を行った。その結果、重決定係数(または、寄与率ともいい R^2 で示す)は 0.030 と極めて低かった。両変数のピアソンの相関係数も $r=.173(n.s.)$ で、有意ではない。Y軸に日本語能力の総合得点を取り、X軸に Can-do Scale の総合得点を取って、34名の被調査者を図1のようにプロットして、日本語能力を日本語版 Can-do Scale で予測する線形近似直線を描いた。

図1の線形近似直線は、最小2乗法(least squares)で両変数の得点の残差(または誤差)が最小になるように計算したものであり、両変数のデータの最も近い距離に位置する直線である。プロットから過小評価だと思われるのは、図1に示したように4名で、全員が中国または台湾出身である。一方、Taylor & Brown (1994)が報告しているように、本調査でも極端に過大評価しているのではないかと思われる被調査者が3名みられる。これら3名は、日本語能力が近似直線よりも30点以上も下回っていた。その1名はアメリカの出身で、日本語能力の合計得点が23点と極めて低いにも拘らず、自己評価は49点であった。また、他の1名はオーストラリア出身で、日本語能力が35点で自己評定が66点であった。もう1名は韓国出身で、日本語能力が29点で、自己評価が63点であった。図1に示したように過大評価と思われる被調査者が2名いたが、いずれも中国出身者であった。これらの自己の過大・過小評価は、やはり性格、文化的背景、言語に対する不安などからくるのであろうが、本研究のデータは、これらの条件を統制していないため、傾向を示唆するくらいしか語れない。

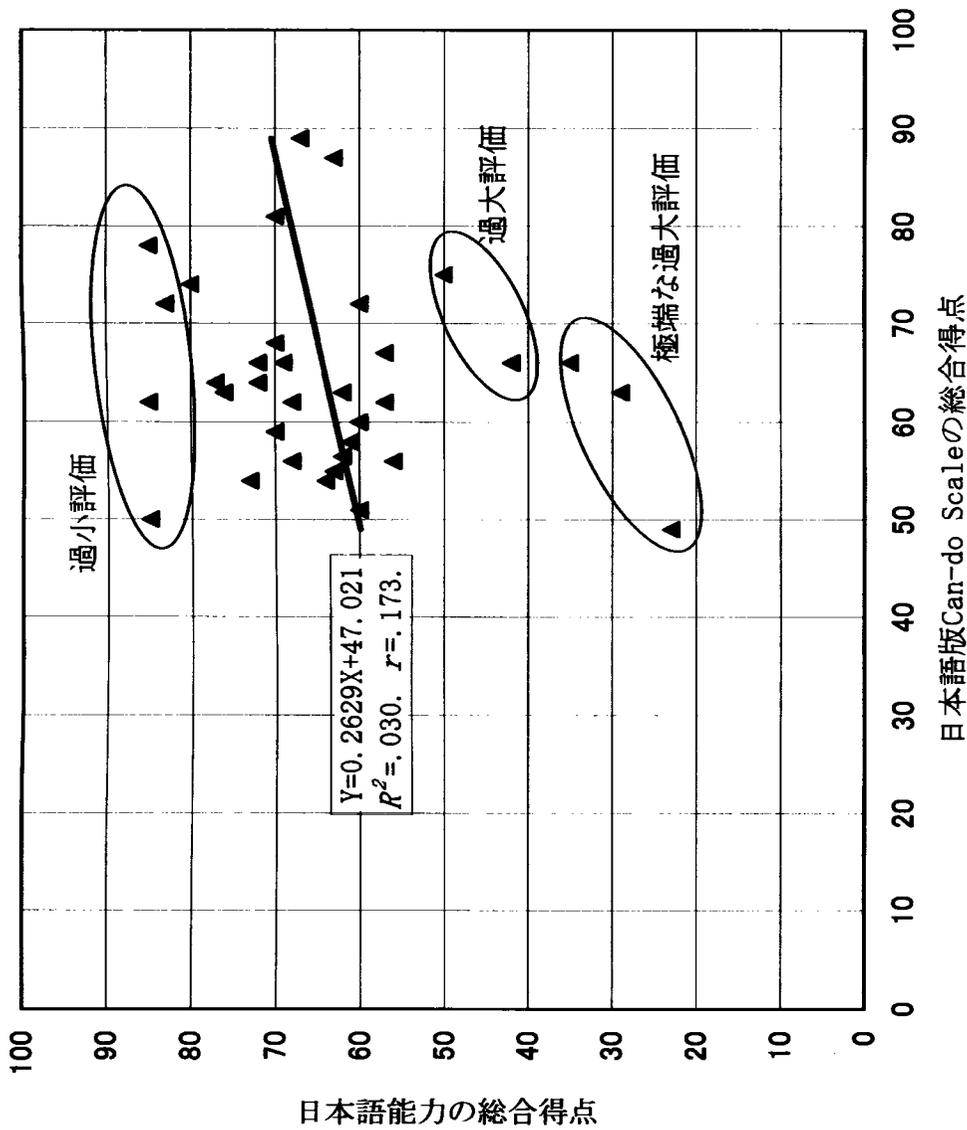


図1 プロットイングでみる日本語能力の過大および過小評価

6. 総合考察

当初、日本語版 Can-do Scale は、適切な質問紙を作成すれば、学習者は正確に日本語能力を自己評価できると考えられてきた。外国語技能に関する先行研究(e. g., Blanche & Merino, 1989; Leblanc & Painchaud, 1985)では、言語技能と自己評価に全体としてかなりの関連があるとされてきた。しかし、本研究は、質問紙形式で測定した日本語版 Can-do Scale が、テスト形式で測定した日本語能力とほとんど関係のない指標であるという結果を示した。実際、自己評価と言語能力とが一致しないという報告は少なくない(e. g., Dieten, 1990; Gardner, Lalonde, Moorcroft & Evers, 1987; Kraemer & Zisenwine, 1989; MacIntyre, Noels & Clément, 1997)。とりわけ、MacIntyre et al. (1997)では、言語能力テストの内容と Can-do Scale の自己評定項目を合わせて測定しているが、結果にかなりのズレがみられたことは、実際の言語能力と自己評定との違いを顕著に示している。このような他言語についての先行研究を加味して考えると、日本語版の Can-do Scale の自己評価であっても、日本語能力を測定するには「本質的に妥当でない(inherently unreliable)」(Oskarson, 1984, 1990; Holec, 1988)と結論づけるのが適切ではないかと思われる。少なくとも、学生の選抜、授業の成績、履修の証明など、直接日本語能力を評価するような目的には使うべきではないといえよう。言い換えれば、学習者間の日本語能力の比較評価には使うべきではないであろう。

確かに調査1では、日本語版 Can-do Scale は、質問紙調査として高い信頼性を示した。しかし、日本語能力を構成する下位カテゴリーとして、「聞く」「話す」「書く」の3技能が一つの因子となり、これらは本研究の質問項目では区別できない技能であった。「読む」技能については、単独で一つの因子となった。この結果は、たとえ20項目の質問が同じ内容を一貫して測定しているとしても、その測定結果は、質問の意図に反して、4つの言語技能を区別しうるようなものではないことを意味している。先行研究の MacIntyre, et al. (1997)の26項目からなる英語版 Can-do Scale でも、「読む」能力が他の「聞く」「話す」「書く」能力と異なる要素を含んでいることが示されており、英語および日本語で類似した因子構造を示した。

調査2では、日本語版 Can-do Scale は日本語能力を測定するという意味での妥当性が低いことが分かった。日本語版 Can-do Scale で測定した4技能と日本語能力テストの得点との相関が「聞く」「話す」「読む」「書く」のすべての指標できわめて低かった。さらに、日本語能力テストの結果を日本語版 Can-do Scale の4技能で予測する重回帰分析を行った。その結果、日本語版 Can-do Scale で尋ねた日本語能力に対する自己評価は、実際の日本語能力(少なくとも本研究でテストした文字言語を媒介とした総合日本語能力)

を予測できなかった。これには、日本語版 Can-do Scale と日本語能力試験とが測定している内容が違っていることが予想される。

日本語版 Can-do Scale が日本語能力を具現化していないという本研究の結果は、個人の社会・文化的背景、性格要因、言語に対する学習・表現不安など多様な要因によって、日本語学習者が自己を過大あるいは過小に評価してしまうことから生じたのではないかと思われる。本研究の調査はこれらの要因について議論するだけのデータではないが、図1が示すように、ある程度の性格や文化要因が影響している傾向がみられた。第二言語に対する不安についての研究はすでに進んでおり、第二言語の成績や課題遂行得点との間に負の相関がみられることが報告されている (e. g., Horwitz, 1986; Horwitz, Horwitz & Cope, 1986; MacIntyre & Gardner, 1989)。近年は、日本語の第二言語不安についての研究も行われている (e. g., 元田, 1999, 2000; 小河原, 2001)。今後、こうした要因を含んだ日本語能力の自己評価に関する研究が進めば、Can-do Scale が測定している内容についての議論が深められ、言語能力試験の代替としてではなく、教授・学習過程での学習目標や成果、自己フィードバックとして利用する可能性も明らかになるのではないかと期待される。

引用文献

- J. D. ブラウン著・和田稔訳 (1999) 『言語テストの基礎知識』大修館書店。
- 元田静 (1999) 「初級日本語学習者の第二言語不安についての基礎的調査」『日本教科教育学会誌』21(4), 45-52.
- 元田静 (2000) 「日本語不安尺度の作成とその検討—目標言語使用環境における第二言語不安の測定—」『教育心理学研究』48, 422-432.
- 日本語教育学会 (1999) 『Can-do-statements 調査報告—日本語能力試験の妥当性を検証する試み』日本語教育学会。
- 日本語教育学会 (1991) 『日本語テストハンドブック』大修館書店。
- 小河原義朗 (2001) 「外国人日本語学習者の日本語発音不安尺度作成の試み—タイ人大学生の場合」『世界の日本語教育』11, 39-53.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implication for teachers and researchers. *Language Learning*, 39, 313-340.
- Dieten, A. J. (1990). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6, 30-46.

- Ferguson, N. (1978). Self-assessment of listening comprehension. *International Review of Applied Linguistics in Language Learning*, 16, 149-156.
- Gardner, R. C., Lalonde, R. M., Moorcroft, R., & Evers, F. T. (1987). Second language attrition: The role of motivation and use. *Journal of Language and Social Psychology*, 6, 29-47.
- Holec, H. ed. (1988). *Autonomy and foreign language learning*. Strasbourg: Council of Europe.
- Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *TESOL Quarterly*, 20, 559-562.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Learning*, 70, 125-132.
- Kraemer, R., & Zisenwine, D. (1989). Changes in attitude toward learning Hebrew in a South African setting. *Language Learning*, 39, 1-14.
- Leblanc, R., & Painchaud, G. (1985). Self-assessment as a second-language placement instrument. *TESOL Quarterly*, 19, 673-686.
- MacIntyre, P.D., & Gardner, R. C. (1989). Anxiety and second-language learning. *Language Learning*, 39, 251-275.
- MacIntyre, P.D., Noels, K.A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, 47, 265-287.
- Oskarsson, M. (1984). *Approaches to self-assessment in foreign language learning*. London: Pergamon.
- Oskarsson, M. (1990). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6, 1-13.
- Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin*, 116, 21-27.