

SPOTをクオリファイングテストとして活用するための基礎研究

竹内 明弘

国際大学

要旨

この論文はSPOTを国際大学のクオリファイングテストとして何らかの形で使用する可能性を探るための基礎研究である。先行研究として上級コースを対象に行われた竹内(2002)の分析を初級、中級レベルにも行ったものである。初級と中級の成績とSPOTとは相関がかなり見られ、このことはSPOTがクオリファイングテストとして組み込むことが可能なことを示唆しているが、上級での使用には問題がある。

キーワード：SPOT、プレースメントテスト、クオリファイングテスト、ピアソン相関、t検定、ANOVA

1 はじめに

筑波大学の小林ら(小林、フォード丹羽、山本 1995、1996)によって開発された、非カリキュラム依存型のSPOT (Simple Performane-Oriented Test) は簡便さと、大量の学習者を短時間で効率よく正確に弁別できるプレースメントテストとして日本語教育機関で利用されて久しい。他のテストを併用することによって、特定の教育機関のコースにより相応しいプレースメントを行う工夫をしているケースも見受けられる。(Toyoda・Hashimoto 2002)

国際大学ではカリキュラムに密接に結びついた、到達度テスト的性格を持つクオリファイングテストを学生に課しており、本稿もSPOTを何らかの形で工夫してクオリファイングテストに使用する可能性を探るものである。

竹内(2002)は国際大学の上級学習者を対象に行った研究ではSPOTとコースの成績との相関があまりなく、縦断的測定では上級学習者の成績の推移がSPOTに現れてこないとして上級のクオリファイングテストには使えると結論できないという結果を出した。本研究はSPOTを初級と中級のレベルにも実施した2年間のまとめである。

なお、エレメンタリーの学習者の成績項目とSPOTとの相関は本紀要の木村(forthcoming)の論文を参照されたい。この論文と併読することによりSPOTと学習者の関係の全体像が理解できると思われる。

本研究でSPOTを実施した日本語のレベル、時期、バージョンを表1に示す。

コースの実名と略称、また SPOT のバージョンと実施時期との関係を参照されたい。

02 年秋学期以前	02 秋学期	03 冬学期	03 春学期
02 ベーシック3 (02B3)	ベーシック 1(02B1)	ベーシック 2(03B2)	ベーシック 3(03B3)
SPOT-B			SPOT-B
02 ベーシック 6 (02B6)	ベーシック 4(02B4)	ベーシック 5(03B5)	ベーシック 6(03B6)
SPOT-B			SPOT-B
02 エレメンタリー-3 (02E3)	エレメンタリー-1(02E1)	エレメンタリー 2(03E2)	エレメンタリー 3(03E3)
SPOT-B		SPOT-B (t1)	SPOT-B (t2)
00 中級1 (00Int1)	中級 1(02Int1)	中級 2(03Int2)	中級 3(03Int3)
SPOT-B	SPOT-A(t1)&B		SPOT-A (t2)
01 上級 1 (01A1)	上級 1(02A1)	上級 2(03A2)	上級 3(03A3)
SPOT-A	SPOT-A (t1)		SPOT-A (t2)

(表 1)

2 構成

分析は以下の 5 部に分ける。

i) 異なるコース間の差異を ANOVA で測定

a) 同時期 (2003 年春学期) に行われた異なる 3 つの初級コース間の差異 :

03E3-03B3-03B6 間

b) 2 年にまたがった春学期の 3 種類の初級コース間の差異 :

02B3-02E3-03B3-03E3-03B6 間

ii) 同一コースの初め (t1) と終わり (t2) で SPOT の成績の縦断的変化を
ペアワイズ t 検定で測定 :

a) 02Int.1-03Int.3 間

b) 02A1-03A3 間

iii) 年度の異なる同一コースの学習者間の差異を t 検定で測定 :

a) 00Int.1-02Int.1 間

b) 01A1-02A1 間

iv) 各コースの SPOT 実施学期の成績項目と SPOT との関連をピアソンの相関で測定 :

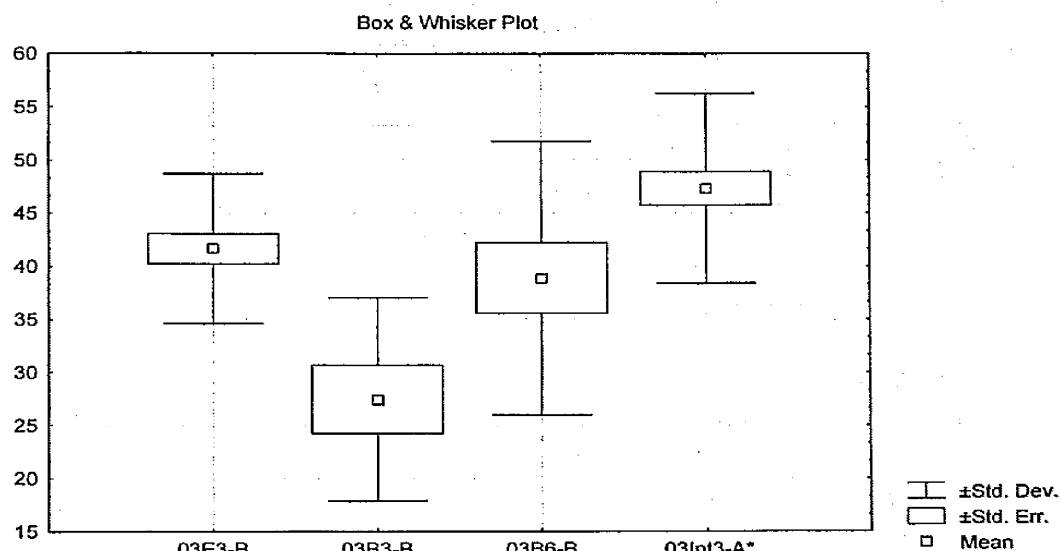
03B3、03B6、00Int.1、02Int.1 (SPOT バージョン A と B)、03Int.3、
01A1、02A1、03A3

v) 3 年にまたがる同 1 コースの開始と終了時期での SPOT の成績を ANOVA で比較する : 00~03 の A1 と A3

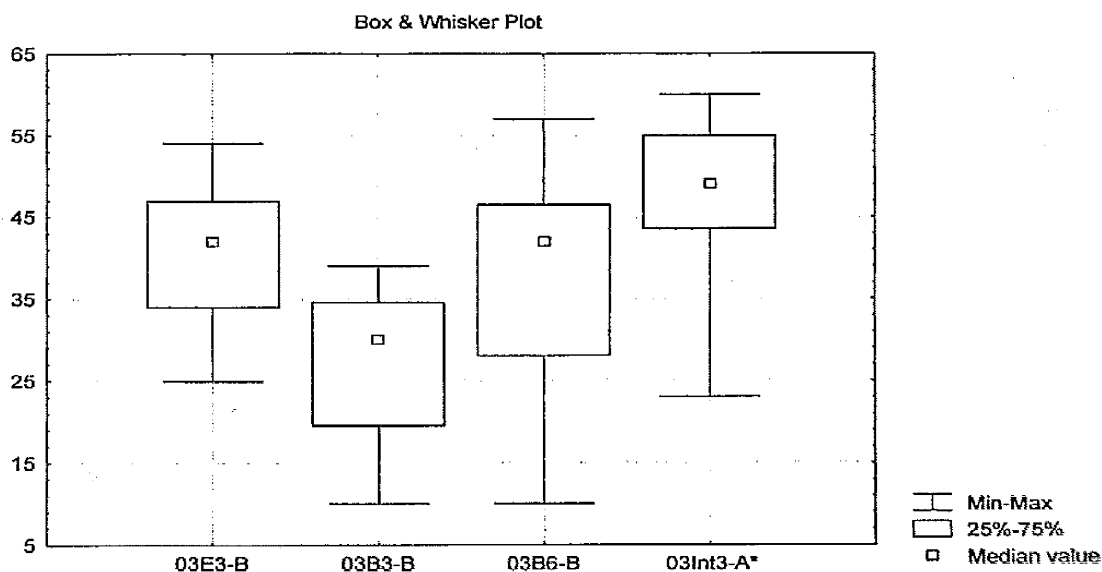
3 分析と解釈

i) 異なるコース間の差異をANOVAで測定

- a) SPOT-Bを出口テストとして、異なるコース間の差異をANOVAで測定: ベーシック3-ベーシック6間、ベーシック3-エレメンタリー3間、ベーシック6-エレメンタリー3間



(グラフ1)



バージョン A (65点満点) を私用した 03Int.3-A は参考値 (グラフ 2)

One-way Analysis of Variance

SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error
1130.07841	2	565.039203	3769.82857	40	94.2457143
F	p				
5.99538352	0.005281**				

Scheffe test	03E3	03B3	03B6
	M=42.810	M=28.143	M=38.867
03E3			
03B3	0.0052966		
03B6	0.4921584	0.065976	

Scheffe test

03E3 vs 03B3 $p=0.0052966 < .01$ **

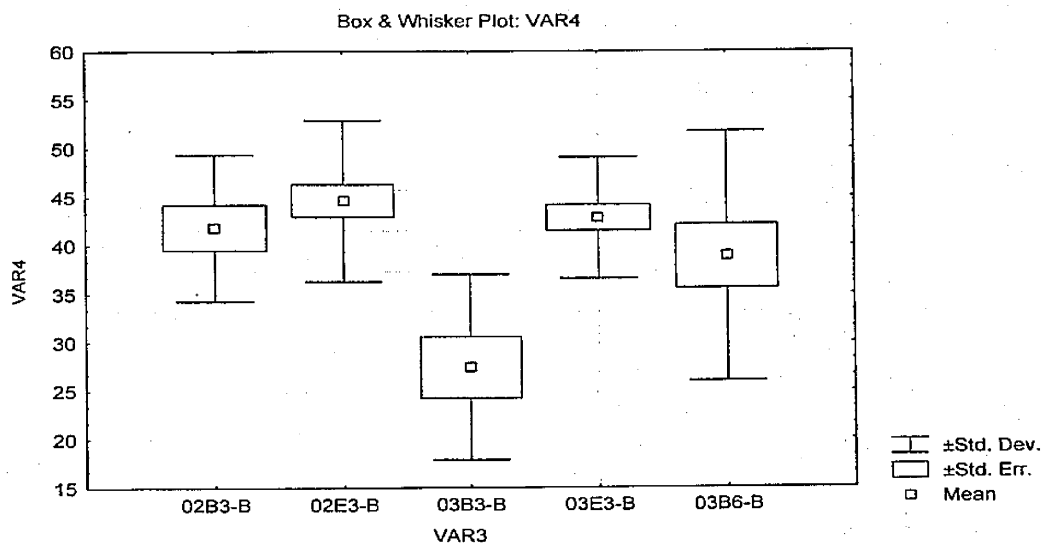
03B3 vs 03B6 $p=0.065976$

03E3 vs 03B6 $p=0.4921584$

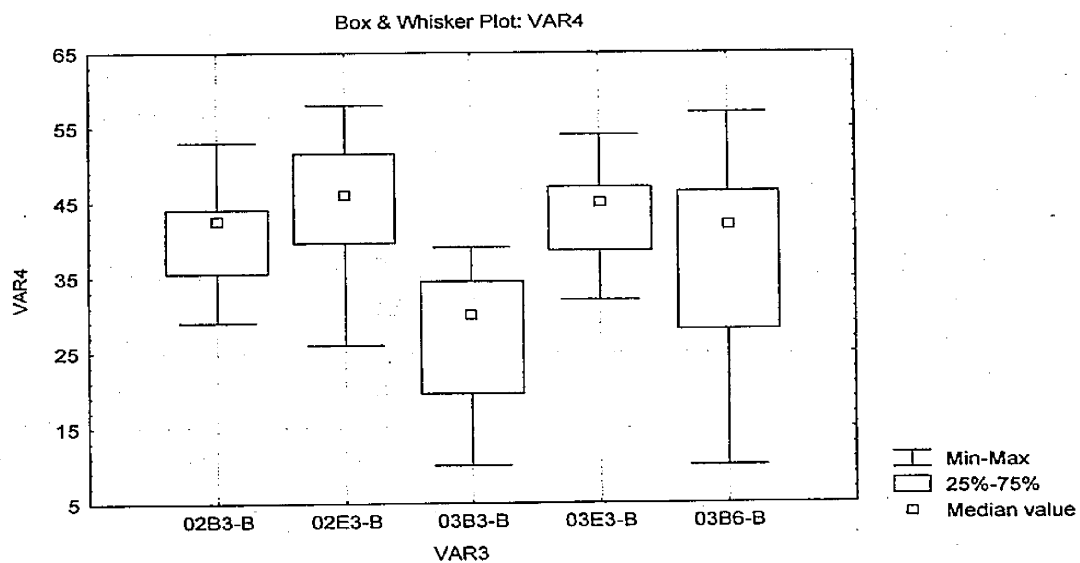
以下 ANOVA と t 検定での確率は 5 % 以下を * で、1 % 以下を ** で表すこととする。

03年エレメンタリー3と03年ベーシック3との間には1%水準で有意差が見られた。グラフ2から分かるようにベーシック6の最高値と最低値はエレメンタリー3の最大値からベーシック3の最低値にまたがっている。これがベーシック6とその他2つのコースとの有意差が認められない理由であると考えられる。

b) ANOVA による 02B3-02E3-03B3-03E3-03B6 間の差異の測定



(グラフ 3)



(グラフ 4)

One-way Analysis of Variance

SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error
2117.1563	4	529.28909	6000.7937	75	80.010582
F	p				
6.6152386	0.00013 < .01**				

Scheffe test

	02B3	02E3	03B3	03E3	03B6
	M=41.800	M=44.600	M=27.444	M=42.810	M=38.867
02B3					
02E3	0.9506015				
03B3	0.0219266	0.0002694			
03E3	0.999073	0.9770991	0.0020922		
03B6	0.9572322	0.432912	0.0672	0.7901045	

03B3vs02B3 p=0.021927 < .05 *

03B3vs02E3 p=0.0002694 < .01 **

03E3vs03B3 p=0.0020922 < .01 **

03年ベーシック3と02年ベーシック3との間には5%水準の、03年ベーシック3と02年エレメンタリー3の間と、03年エレメンタリー3と03年ベーシック3の間には1%水準の有意差が見られた。

日本語コースの作成者がコースの成り立ちから妥当だと予想した差異は以下のとおりである。

1) エレメンタリー3であれば、ベーシック3より成績が上である。

2) エレメンタリー3とベーシック6ははっきりとして差異はみられない。

1) が予想に反したのは02年のベーシック3と03年のエレメンタリー3との間に有意差が認められなかったことが原因であるが、グラフからも平均値、中央値、標準偏差でこの2集団には差がない。

また2) は予想の通りであるが、分布が広範にわたっていることが原因と見られる。

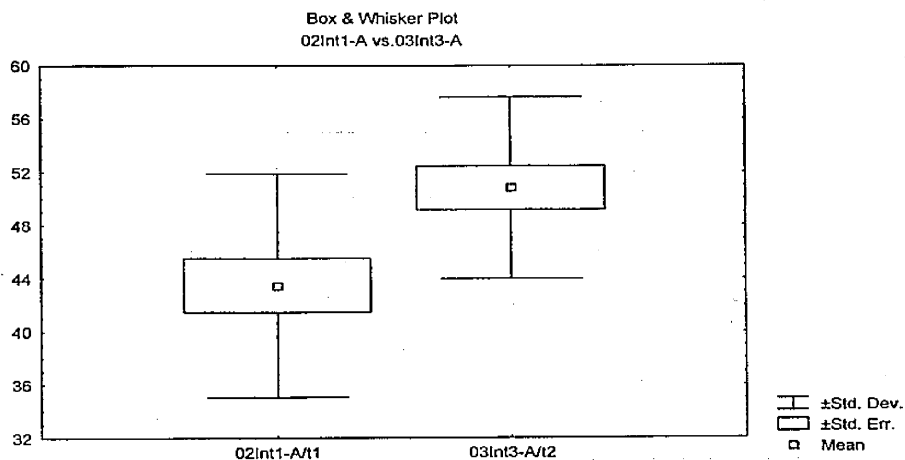
有意差が認められた2集団間の特徴はコース間ではなく、コースを履修した年にあると言えよう。

つまり、02年の履修者は総体的に03年の履修者より成績が上であったということである。

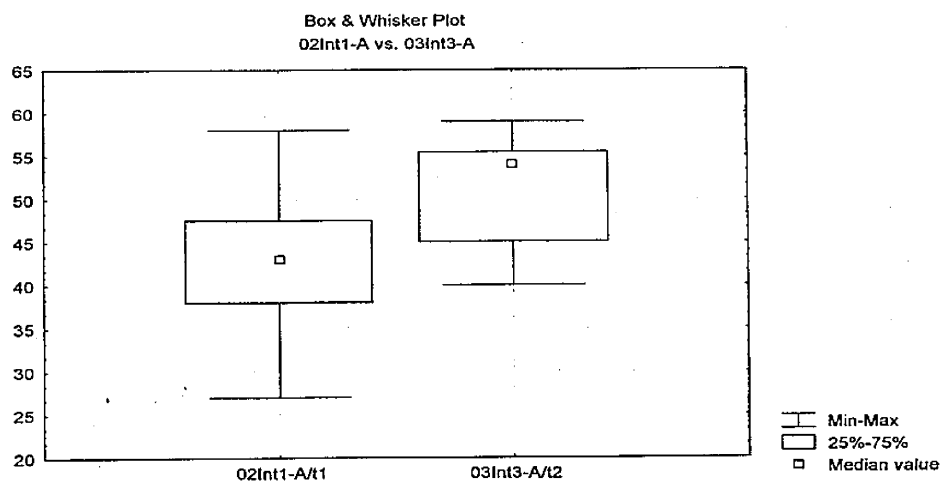
ベーシック6は初級1年間終了程度の学習者を対象にしているというコースの性格と一週間に2回(一回90分)という少ない時間数であることから、言語能力のプロファイルが極めて多様性に富む学習者が集まった。それゆえ SPOT の結果もばらつきの大きいものとなったと考えられる。

ii) 同一コースの初め(t1)と終わり(t2)で SPOT の成績の縦断的变化をペアワイズt検定で測定:

a) 02 年中級 1-03 年中級3間



(グラフ 5)



(グラフ 6)

T-test for Dependent Samples

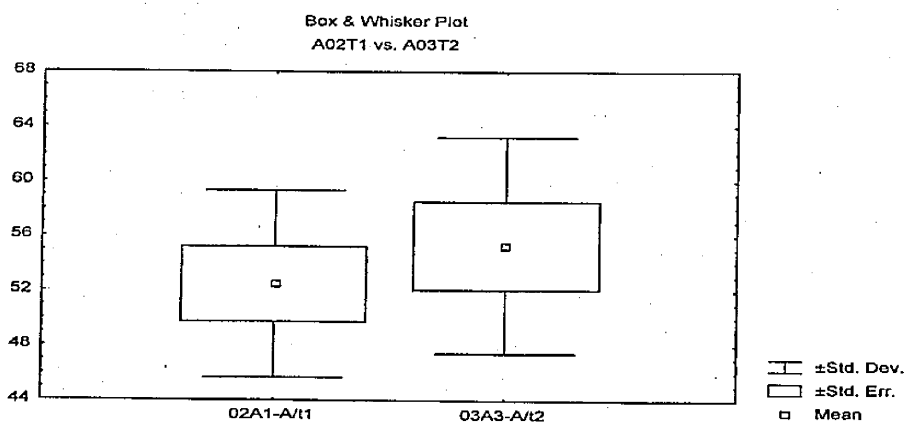
	Mean	Std.Dv.	N	Diff.	Std.Dv.Diff.
T1	43.470588	8.4047431			
T2	50.823529	6.8304035	17	-7.3529412	4.7950649

t	df	p
-6.322532	16	.000010139<.01 **

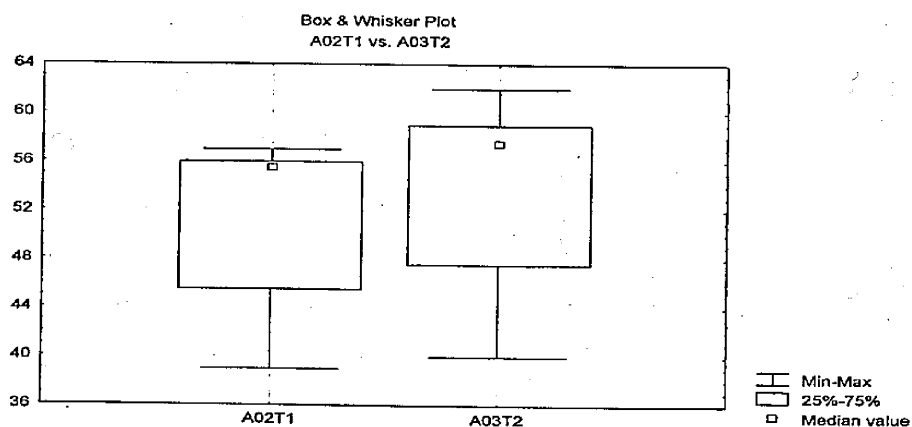
ペアワイズt検定で縦断的に測定した結果、中級1と中級3の間には1%の水準で有意差が認められた。グラフ5から全体に成績が向上していること、グラフ6からは特に最低値が上がっていることが

主因と考えられる。

b) 02 年上級1ー03 年上級3間



(グラフ7)



(グラフ8)

T-test for Dependent Samples

	Mean	Std.Dv.	N	Diff.	Std.Dv.Diff.
T1	52.5	6.8337398			
T2	55.333333	7.8909231	6	-2.833333	2.31660671

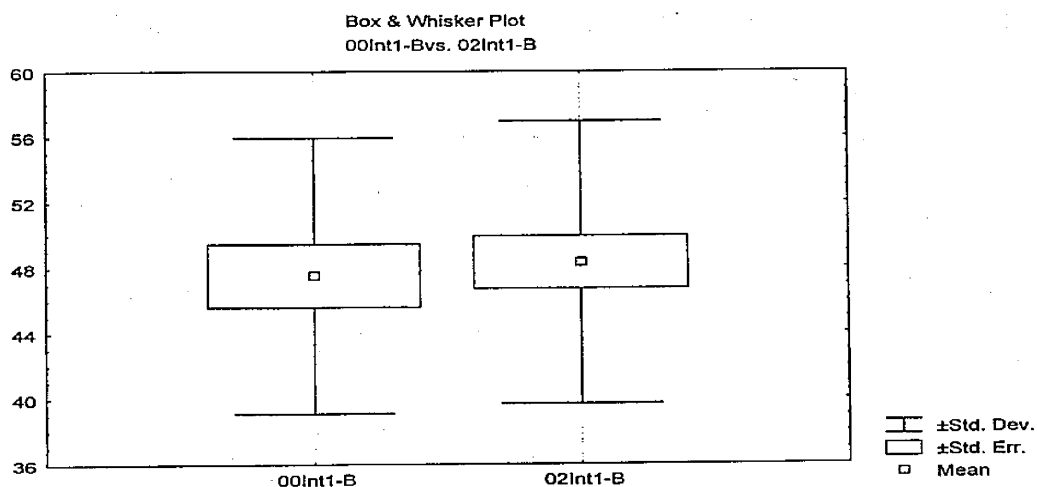
t	df	p
-2.995856	5	0.03024 < .05 *

ペアワイズt検定で縦断的に測定した結果、02年上級1と03年上級3の間には 5%の水準で有意差が認められた。グラフ 7 から全体に成績が向上していること、グラフ 8 から、最低値はさほどかわら

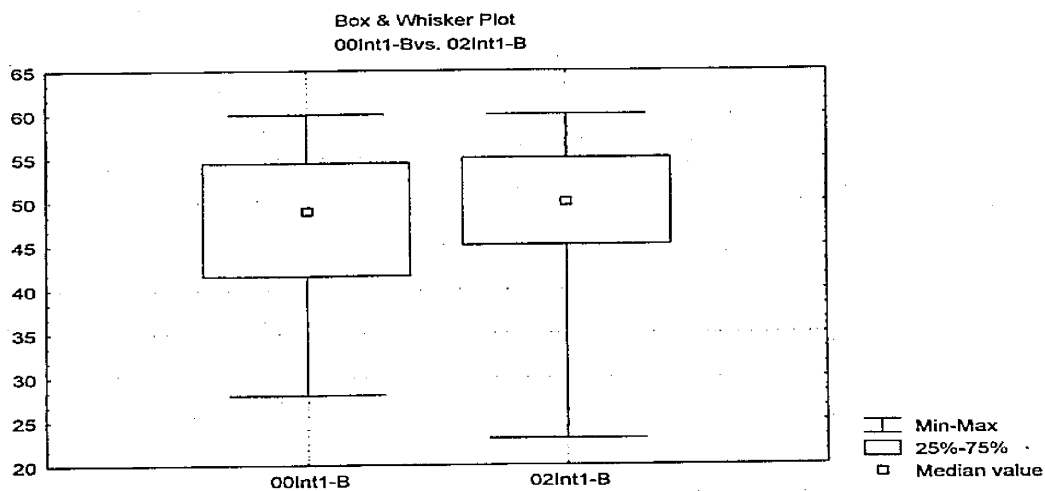
ないが、最高値が上がったことが原因であろう。中級が最低値が伸びたことによって有意差が生じたのと対照的である。

iii) 異なる年度の同一コースの学習者間の差異を t 検定で測定:

a) 00 中級1-02 中級1 間



(グラフ9)



(グラフ10)

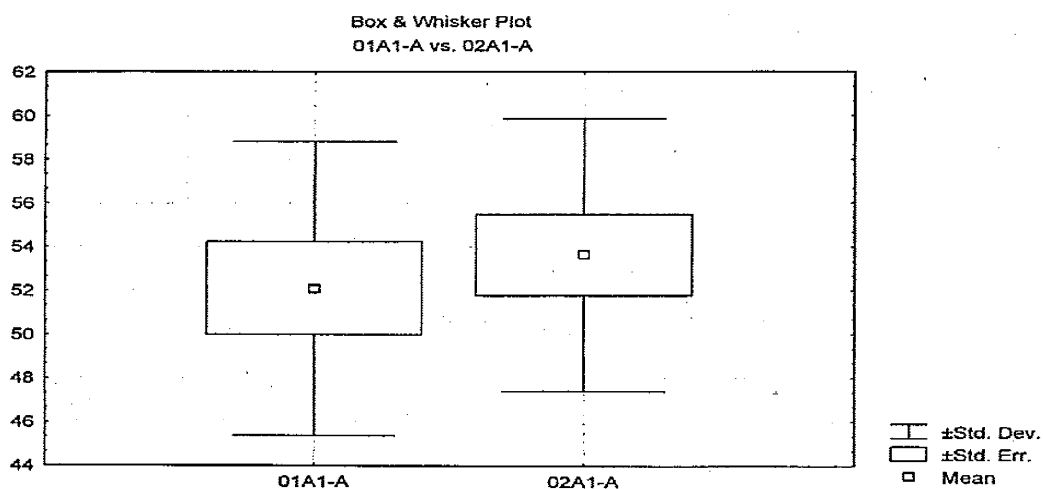
T-test for Independent Samples

Mean 00Int1	Mean 02Int1	t-value	df	p
47.526316	48.321429	-0.312911	45	0.75579424

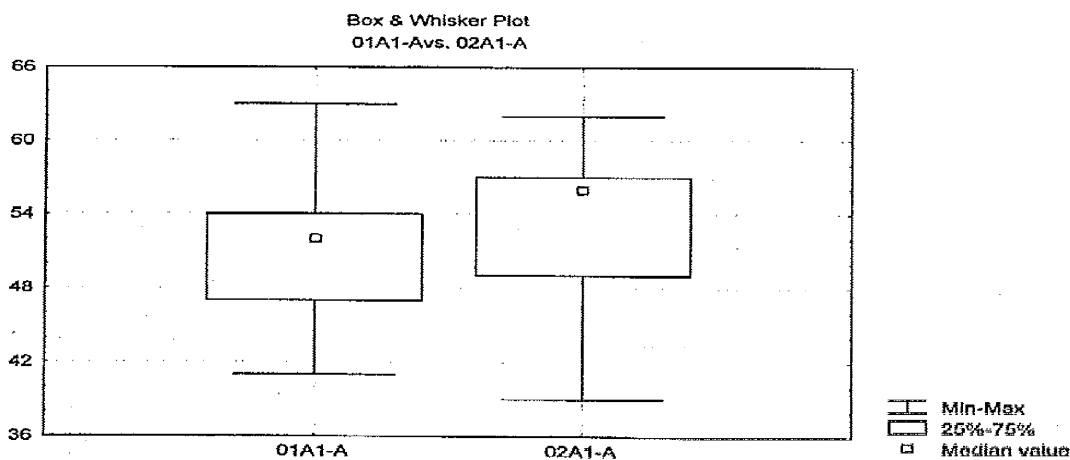
Valid N 00Int1	Valid N 02Int1	SD 00Int1	SD 03Int1	F-ratio variances	p variances
19	28	8.4219846	8.6326408	1.05065094	0.9326886

t検定の結果中級1を履修した00年と02年の2つの集団間には有意差が認められなかった。この2集団にはクオリファイングテストを通して中級1の履修を始めた段階で差がなかったと言えよう。

b) 01 上級1-02 上級1間



(グラフ11)



(グラフ12)

Mean 01A1	Mean 02A1	t-value	df	p
53.636364	52.1	0.5448124	19	0.59222049

Valid N 01A1	Valid N 02A1	SD 01A1	SD02A1	F-ratio variances	P variances
11	10	6.21728	6.707375708	1.1638706	0.8109295

t検定の結果2つの集団間には有意差が認められなかった。この2集団にはクオリファイングテストを
通って上級1の履修を始めた段階で差がなかったと言えよう。

iv) 同一コースの SPOT 実施学期の成績コンポーネントと SPOT との関連をピアソンの相関で測定：
03 ベーシック3、03 ベーシック6、00 中級1、02 中級1 (SPOT バージョン A と B)、03 中級3、01
上級1、02 上級1、03 上級3

はじめにピアソンの相関係数の解釈の基準を岩淵(1997)から引用する。

.00～± .20 ほとんど相関がない
±.20～± .40 低い(弱い)相関がある
±.40～± .70 かなり(比較的強い)相関がある
±.70～±1.00 高い(強い)相関がある (岩淵1997 p.119)

上記の解釈に則り、±.40 以上を相関があるとして、数字をイタリック・ボールドで示す。低い(弱
い)相関があるとなっている±.20～±.40 の値については母集団の小ささから考慮に入れないことと
する。

03 ベーシック3

	中間テスト	中間テスト	中間テスト	中間テスト	中間テスト
ピアソン相関 N=9	聞く	漢字	文法表現	作文	合計
SPOTB ひらがな版	0.0219516	<i>0.7533</i>	-0.145401	<i>0.450643</i>	0.2395431

Q	発表	会話	平常	達成	宿題	B3 総合
0.0453421	<i>0.538829</i>	0.16788	<i>0.54199</i>	0.2845689	-0.152498	0.2831477

強い相関があったのは中間テストの漢字で、比較的強い相関がみられた項目は中間テストの作文、
発表、平常であった。

03ベーシック2と SPOT (03ベーシック3の時に実施)との相関

03ベーシック3は SPOT を行っているときにかなり緊張を強いられていたとの報告があった。このことは試験の強いストレスで受験者の能力が退行していた可能性が考えられるので、一つ前の学期の02ベーシック2との相関も調べてみた。

	B2mid				
ピアソン相関 N=9	聞く	語・文法	表現会話	作文	合計
B3B ひらがな版	-0.233492	0.3883553	0.1316526	0.637465	0.413618

B2-final					
聞く	活用語彙	構文	表現会話	作文	合計
0.2745782	0.2235295	0.08103	0.0240838	-0.323536	0.1213858

					B3 総計
Q	発表	会話	宿題	平常	
-0.087505	0.444761	0.584528	-0.51163	0.2644038	0.1955308

比較的強い相関があった項目は中間テストの作文と合計、発表、会話で、宿題とは負の相関があった。これをもって学習者が一つ前の学期のレベルに退行していたというのは難しい。

ベーシック6

ピアソン相関	期末試験				クイズ	発表	会話
	文法	会話表現	作文	期末合計			
B6B N=13	0.506702	0.732592	0.484564	0.644279	0.677652	0.564977	0.51206

平常			宿・参・進	総計
宿題	参加	進歩	平均	
0.60117	0.628872	0.801316	0.763783	0.690089

ベーシック6は全項目と何らかの相関が認められた。強い相関があった項目は期末試験の会話表現、平常点進歩と平均であった。比較的強い相関があった項目は期末試験の文法と作文と合計、クイズ、発表、会話、平常点の宿題と参加、総計であった。言語能力の下位技能をレーダーチャートに表して考えた場合、全項目と相関があったということは以下のことを示唆している。ある学習者のレーダーチャートはその下位に位置する学習者のレーダーチャートより全ての面で上回っており、成績の序列の中で隣接する学習者間の能力差が画然としている。

00 中級1

ピアソン相関	中間	期末	クイズ	漢字	出席
00Int1-B N=19	0.1866237	0.2893696	0.1798772	-0.061919	0.1537264

聞き取り	スピーチ	会話	宿題	平常	00Int1 総計
0.711067	0.597395	0.66444	0.0715572	0.3427641	0.440017

強い相関があったテスト項目は聞き取りであった。比較的強い相関があった項目はスピーチ、会話、総計であった。相関があった項目はいずれも言語の即時的な処理を要求するものであり、SPOT が測っているとされている能力の特徴に一致している。

02 中級1 (SPOT バージョン A と B)

02中級1の時に実施した SPOT バージョン A(t1)と同中級1のサブスキル項目、同 SPOT と03中級3のサブスキル項目との相関をみた。後者は予測性があるかを調べるために分析をした。

ピアソン相関	02int1						
	話す	聞く	構文	漢字	出欠	平常	合計
At1-int1 N=28	0.679735	0.711403	0.446629	0.3482349	-0.152948	-0.063886	0.503656
At1-int3 N=16	0.2460033	0.3913344	-0.019737	-0.076175	-0.286538	-0.097374	0.0615695

強い相関があったサブスキル項目は「聞く」であった。比較的強い相関があったものは「話す」と合計であった。03中級3の項目と相関があったものはなかった。相関があった項目は00中級1の結果と同様に言語の即時的処理を要求されるものであった。

03 中級1と同学期に実施した SPOT バージョン B

ピアソン相関	02int1						
	話す	聞く	構文	漢字	出欠	平常	合計
B-int1 N=28	0.687811	0.755347	0.557806	0.409817	-0.094226	0.0402318	0.598067

強い相関があったサブスキル項目は「聞く」であった。比較的強い相関があったものは「話す」、「構文」、「漢字」と合計であった。相関のあるものは即時的処理を要求するものであった。

SPOT の2つのバージョンを比べると、AよりBの方が該当学期のサブスキル項目の数と係数の高さにおいてより強い相関があるといえる。

03中級3の時に実施した SPOT バージョン A(t2)と同中級3との相関

ピアソン相関	03int3						
	話す	聞く	構文	漢字	出欠	平常	合計
At2-int3 N=21	0.27314	0.71566	0.2518315	0.0115237	-0.077311	-0.044383	0.2830498

かなり強い相関が認められたサブスキル項目は「聞く」であった。

02 中級 1 の時に実施した SPOT バージョン A(t1)と同中級 1 の試験項目との相関

ピアソン相関	02int1					
	話Q	会話1	会話2	発表	聞Q	聞期末
Aint1t1 N=28	0.238929	0.657594	0.617068	0.470416	0.734586	0.572275

構Q	構中間	構期末	漢Q	漢中間	漢期末	宿題	平常	出席
0.2382415	0.38306	0.499085	0.116986	0.2302445	0.550044	-0.12704	0.329393	-0.15039

強い相関があった試験項目は聞くクイズで、比較的強い相関があったものは会話1、会話2、期末の聞くテスト、発表、期末の構文テスト、期末の漢字テストであった。

02 中級 1 の時に実施した SPOT バージョン B と同中級 1 の試験項目との相関

ピアソン相関	02int1					
	話Q	会話1	会話2	発表	聞Q	聞期末
Bint1 N=28	0.3874247	0.47156	0.52772	0.537206	0.778037	0.612002

構Q	構中間	構期末	漢Q	漢中間	漢期末	宿題	平常	出席
0.3598627	0.48312	0.601121	0.275765	0.1959797	0.608257	-0.03383	0.450591	-0.09048

強い相関があった試験項目は聞くクイズ、比較的強い相関があったものは会話1、会話2、発表、期末の聞くクイズ、中間の構文テスト、期末の構文テスト、期末の漢字テスト、平常であった。

03 中級 3 の時に実施した SPOT バージョン A(t2)と同中級 3 の試験項目との相関

	03int3						
	話Q	会話1	会話2	発表	聞Q	聞中間	聞期末
Aint3t2 N=28	0.045813	0.541959	0.2333608	0.2252013	0.726536	0.599779	0.510134

構Q	構中間	構期末	漢Q	漢中間	漢期末	宿題	教室活動	出席
0.25391	0.24436397	0.2333619	-0.003838	-0.005	0.0410635	-0.12843	0.312971	-0.06662

強い相関が認められた項目は聞くクイズで、比較的強い相関があった項目は会話1、中間の聞くテスト、期末の聞くテストであった。

中級は全般にわたって、SPOT が測定するとされている言語の即時的処理を要求する項目との相関があると言える。

01 上級1

01 上級1の時に実施した SPOT バージョン A と同上級1の試験項目との相関

ピアソン相関	Q	各週発表	中間テスト						
			ビデオ	読解	エッセー	表現1	表現2	短文	中間合計
A01Ad1	-0.19444	0.533765	0.360485	0.234286	0.306932	0.432573	-0.01024	0.2689	0.3967

期末発表	期末筆記					
	ビデオ	読解	エッセー	語彙	短文	期末合計
0.1500619	0.3890736	0.442721	0.3030603	0.2295172	0.2757102	0.55157

期末会話	平常	進歩	総合
0.584561	0.263071	0.043649	0.464705

比較的強い相関が認められた項目は各週発表、中間の表現テスト1、期末の読解テスト、期末テストの合計、期末会話、総合であった。

02 上級1

02 上級1の時に実施した SPOT バージョン A と同上級1の試験項目との相関

ピアソン相関	Q	発表	期末発表	期末会話
A02Ad1N=10	-0.46646	-0.57562	-0.093161	0.0533938

中間試験						中間試験
聴解	読解	エッセー	文法	語彙	短文	合計
0.3484421	0.3484421	0.46337	-0.3777	-0.283171	0.1528303	-0.098575

期末					期末試験
聴解	読解	作文	語彙	短文	期末合計
0.2244277	0.443404	0.2231705	-0.40226	0.1331811	-0.089098

進歩	平常	総合
0.072831	0.017931	-0.3377

比較的強い相関が中間エッセー、期末読解にあり、比較的強い負の相関がクイズ、発表、期末の語彙テストに認められた。

03 上級3

03 上級3 の時に実施した SPOT バージョン A と同上級3 の試験項目との相関

	Q	発表
ピアソン相関		
A03Ad3 N=6	-0.37045	-0.31749

中間試験						合計
聴解	読解	作文	語彙	慣用句	短文	
0.194813	0.387923	-0.399946	0.1711181	-0.06173	-0.1734	0.0997213

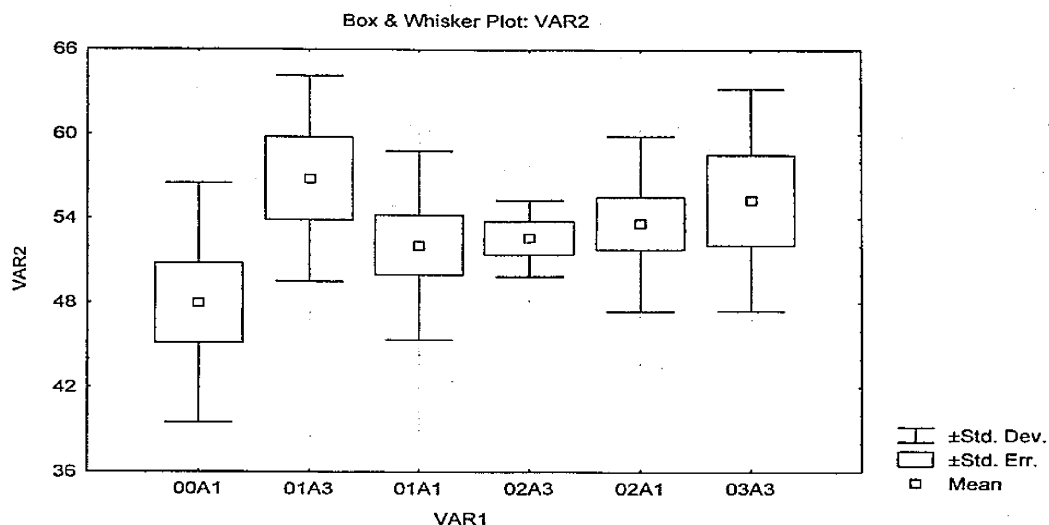
				合計
聴解	読解	作文	語彙	
0.2297388	0.507897	0.3371508	0.668146	0.574676

期末発表	課題提出	期末会話	平常	進歩	総合
0.3988833	-0.95876	-0.115112	-0.07743	-0.13711	-0.1727

比較的強い相関は期末読解、期末語彙、期末合計に、負の強い相関が課題提出に認められた。負の相関が認められた項目は、クイズ、発表、期末の語彙テスト、課題提出であった。これらの項目は言語学習においては比較的時間をかけて行われるものであり、いわゆる地道に努力をするもので、即時的な言語処理とは対極の性質をもつものである。実際課題提出に関しては、「聞く」、「話す」といった即時的処理の苦手な学習者がよく努力しているのが観察された。

上級においては、SPOT との相関のあった項目は必ずしも言語の即時的処理と直結するものではなかった。上級のコースで習得される言語技能と SPOT で測定する能力とは中級のように一致しておらず、乖離があるようである。

v) 同 1 コースの開始と終了時期での成績を ANOVA で比較する: 上級 00~03



(グラフ13)

Analysis of Variance

SS Effect	df Effect	MSEffect	SS Error	df Error	MS Error	F	p
356.1240	5	71.22481	1976.812	41	48.21493	1.477236	.218201

Scheffe Test

	00A1 M=48.000	01A3 M=56.833	01A1 M=52.100	02A3 M=52.600	02A1 M=53.636	03A3 M=55.333
00A1						
01A3	.342639					
01A1	.891745	.880238				
02A3	.920243	.959490	.999998			
02A1	.661381	.974130	.998252	.999909		
03A3	.554057	.999597	.974804	.994292	.998628	

ポストホックテストの結果どの2項目間にも有意差が認められなかった。このことは上級コースでは SPOT のバージョン A の頭打ち効果を示唆しているのかも知れない。

結論

初級レベル: ベーシックに関しては、比較的高い相関が認められたことにより、SPOT はベーシック

3についてはある程度、ベーシック6はかなりの正確さで測定することができると言える。

中級レベル: 中級では学習者の能力の推移を的確に捉えていた。SPOT のバージョンに関しては B より A の方が正確に測定できるようである。

上級レベル: 上級コースが学習者に要求する内容と SPOT で測定しているものとの間に、また学習者によって習得された能力と SPOT で測定しているものとの間には乖離が見られる。

以上のことから、ベーシックの2つのコースには SPOT のバージョン B、中級にはバージョン A をクオリファイングテストに組み込むことが可能だと言える。どのように組み込むかはクオリファイングテストの目的と性質を考慮に入れた上で考えられなければならない。

参考文献

岩淵千明 (1997) 『あなたもできる データの処理と解析』 福村出版

小林典子・フォード丹羽順子・山本啓史 (1995) 「『日本語能力簡易試験 (SPOT)』の得点分布傾向」『筑波大学留学生センター日本語教育論集』第 10 号: 107-120

小林典子・フォード丹羽順子・山本啓史 (1996) 「文法項目聴取能力と音声環境—SPOT (Simple Performance-Oriented Test) の空欄位置に関する一考察」『筑波大学留学生センター日本語教育論集』第 12 号: 125-137

竹内明弘 (2002) 「日本語上級コースの実践報告と SPOT の縦断的測定の結果」 *Working Papers* Vol.12 2002, International University of Japan

Toyoda, E. and Y. Hashimoto. (2002) Improving a Placement Test Battery: What Can Test Analysis Programs Tell Us? *ASAA e-journal of asian linguistics & language teaching*