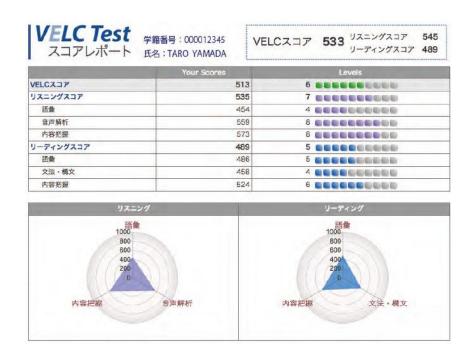
大学生のための 新しい英語テストの開発

靜 哲人
VELC研究会
大東文化大学

VELC テストとは

 Visualizing English Language Competency 日本人大学生の英語力を可視化する



VELCテストの目指したもの

- 日本人大学生に適している (well-targeted)
- 特定のコースに基づかない熟達度テスト (proficiency)
- クラス分けに使える (placement)
- 学習効果の測定に使える (achievement)
- フィードバックが詳しい (diagnosis)
- フィードバックが速い (prompt feedback)

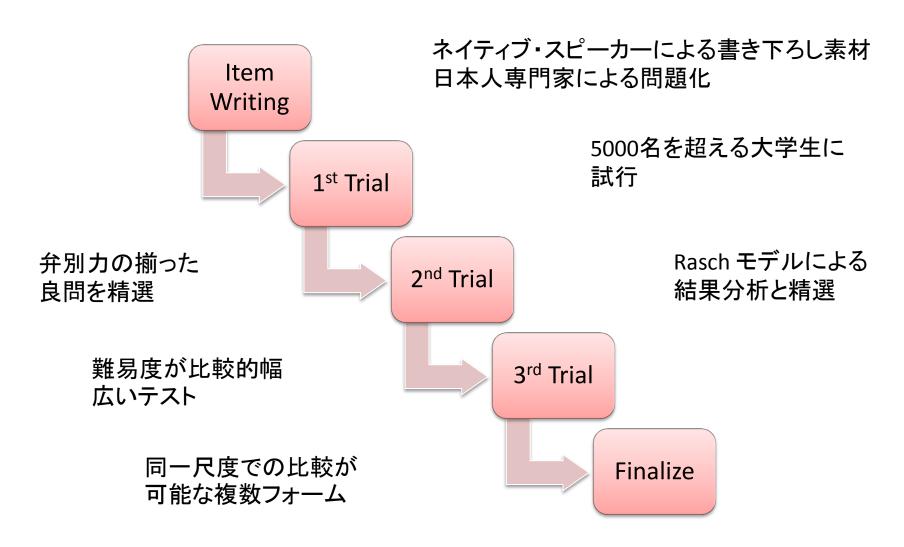
テストの構成

	構成概念名	内容	項目数
Lstng Prt 1	聴解語彙力	リスニングでの語彙サイズ	20
Lstng Prt 2	音声解析力	音の連続を語に切り分ける力	20
Lstng Prt 3	内容把握力	リスニングでの内容を追う力	20
小計			60
Rdg Prt1	読解語彙力	リーディングでの語彙サイズ	20
Rdg Prt 2	構文解析力	語連続をフレーズに切り分ける力	20
Rdg Prt 3	内容把握力	リーディングで内容を追う力	20
小計			60
合計			120

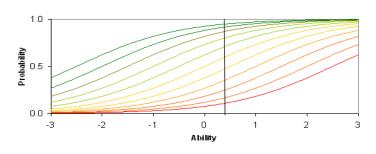
所要時間 約70分 進行はCDで制御

Part	Item Format
L1	日本語の語句を聞いて、それに相当する英単語を、聴覚提示される4つの選択肢から選ぶ。
L2	短い英文を聞いて、指定された位置の語を、視覚提示された4つの選択肢から選ぶ。
L3	ある程度の長さの英文を聞いて、ビープ音によって置換された部分を推測し、視覚提示された4つの選択肢から選 ぶ。
R1	日本語の語句を見て、それに相当する英単語を、視覚提示された4つの選択肢から選ぶ。
R2	1語が欠けた非文を読み、指定された1語を文中のどの位置に戻せば正文となるかを、4つの選択肢から選ぶ。
R3	ある程度の長さ(30~80語程度)の英文に設けられた空所に補充すべき語句を、4つの選択肢から選ぶ。

開発過程



Rasch モデルとは



- 項目応答モデル群 (item response models)の ひとつ
- 最も「厳しい」モデル
- このモデルに適合する問題項目は、弁別力が揃っている
- ・ 別々のテストを受けた受験者同士を同一の 尺度で比較するための理論的根拠



VELCスコア 533 リスニングスコア 545

	Your Scores	Levels
VELCXIP	513	6
リスニングスコア	535	7 000000000
語彙	454	4
音声解析	559	8 000000000
内容把握	573	8
リーディングスコア	489	5
語彙	486	5
文法・構文	458	4
内容把握	524	6 000000000



VELCスコアの意味

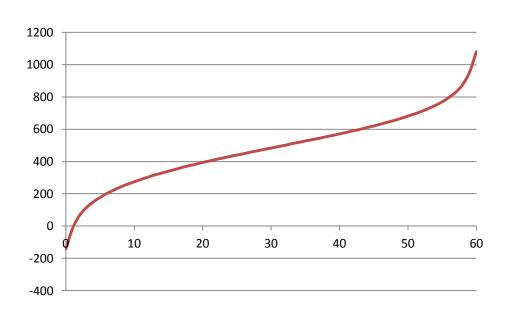
	下からの	上からの
ベルクス	パーセンタイ	パーセンタイ
コア	ルランク	ルランク
250	1%	99%
300	2%	98%
350	7%	93%
400	16%	84%
425	23%	77%
450	31%	69%
475	40%	60%
500	50%	50%
525	60%	40%
550	69%	31%
575	77%	23%
600	84%	16%
650	93%	7%
700	98%	2%
750	99%	1%

平均 500 標準偏差 100

総合、スキル別、セクション別 すべて例えば、600ならば、下から見て84%の位置にいることを 表す。

素点とVELCスコアの関係

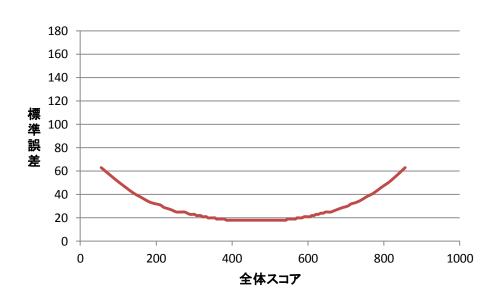
・素点と1対1対応だが、直線的関係でない



VELCスコアの誤差

- 一般的に誤差は、問題項目数が多ければ、能力レベルが中程度であるほど小さくなる。
- 全体スコア(120問に基づく)
- ・ スキル別スコア(60問に基づく)
- ・ サブスキル別スコア(20問に基づく)
- 500に近いスコア
- 100や900に近いスコア

全体スコアの標準誤差



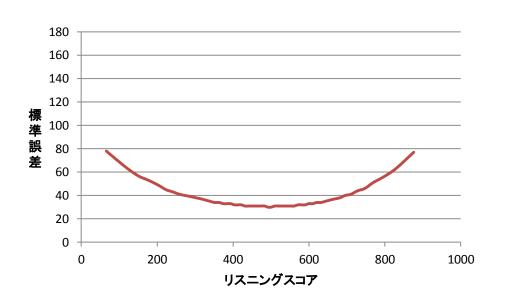
400~600:

200~400; 600~700:

~200; 700~:

おおよそ 18~20 おおよそ 20~30 おおよそ 30~70

セクションスコア(LorR)の標準誤差



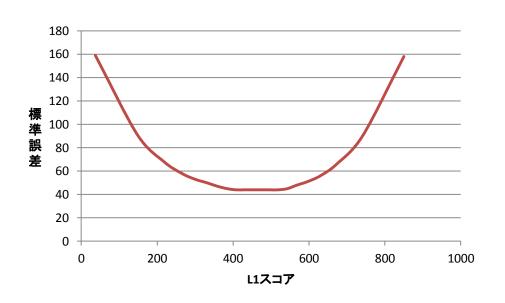
400~600:

200~400; 600~700:

~200; 700~:

おおよそ 30~32 おおよそ 32~45 おおよそ 45~80

パートスコア(L1, L2, ... R3)の標準誤差



400~600:

200~400; 600~700:

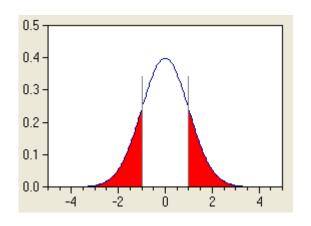
~200; 700~:

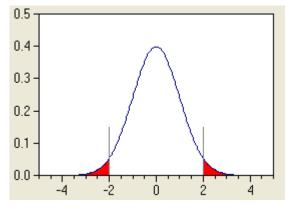
おおよそ 44~50 おおよそ 50~65 おおよそ 65~150

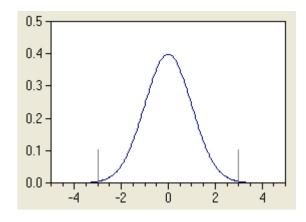
どのくらいスコアが違えば違うのか

- ・標準誤差 = 誤差の標準偏差
- ・誤差は正規分布する
- 正規分布の性質 ± 1 SD --- 約68%

± 2 SD ---約 95.5%



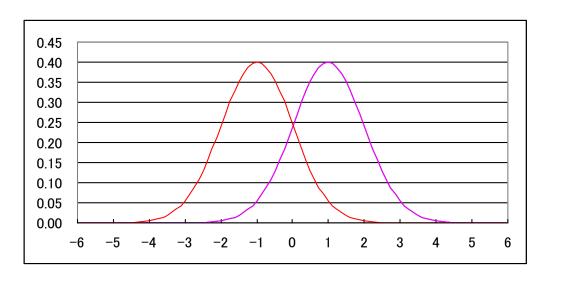


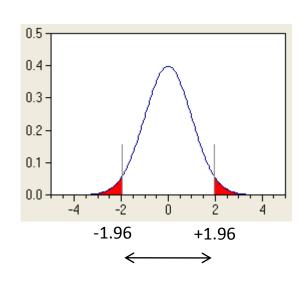


標準誤差の約2倍の差があれば

$$1.96 \times \sqrt{\frac{(SE_1)^2 + (SE_2)^2}{2}}$$

95%違うであろう

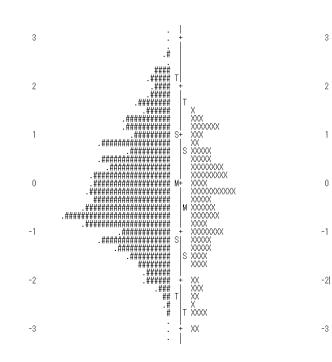




VELCテストの信頼性

・ 幅広い難易度の項目で幅広い受験者に対応

受験者 N = 5583 # が15人を示す



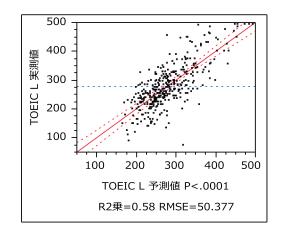
フォームA k = 120

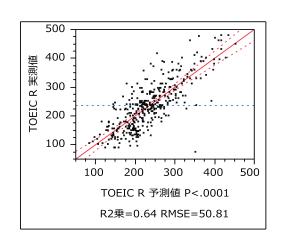
VELCテストの信頼性

- フォームAに含まれる120項目の中で115項目 以上に解答した226名のデータに関して信頼 性係数を算出
- Rasch person reliability = .95
- Rasch item reliability = .95
- この受験者グループに対して非常に高いレベルの信頼性を確認

VELCテストの妥当性

- (TOEICを模しているわけではないが...)
- TOEIC リスニング、リーディンを目標変数、
 VELCのL1, L2, L3, R1, R2, R3を予測変数とした
 ステップワイズ重回帰分析 (N = 375)





VELCテストの妥当性

	TOEIC L	TOEIC R	TOEIC Total
予測変数	L1, L2, L3, R3	L1, L2, L3, R1, R2, R3	左の2つの 単純合計
決定係数	58%	64%	68%
重相関	0.76	0.80	0.82
標準残差	50	51	85

特段、TOEICの予測を目的として開発していない70分、120問のテストで、TOEICの分散の7割近くを説明することが判明した。

知識・スキル別の細分型診断

あなたのスキル別正答率((%)	▲ 閉じる
知識・スキル	あなたの正答率	全国標準正答率
高校レベルの語彙を文脈の助けなしに理解・想起するカ	87%	73
大学基礎レベルの語彙を文脈の助けなしに理解・想起する力	72%	59
大学応用レベルの語彙を文脈の助けなしに理解・想起する力	57%	45
はっきり発音される名詞、動詞、形容詞などを聞き取る力	70% @ 6 6 6 6 6 6 6 6 6 6 6 6	55
弱く発音される前置詞、冠詞、助動詞などを聞き取る力	64%	43
語尾がはっきり発音されない語を聞き取る力	55%	47
あいまいな音で発音される母音を含む名詞、動詞、形容詞などを聞き取る力	80% @@@@@@@@ *	56
語と語がつながって(リンキング)発音されたものを聞き取る力	56%	48
/I/と/r/, /b/と/v/など、日本語にない音の区別を聞き分ける力	80% 0000000000	49
比較的長めの主語・主部を正しく理解する力	58%	51
比較的長めの目的語を正しく理解する能力	59%	51
前置詞で始まる長めの表現を理解する力	52%	53
時や条件、目的などを表す表現(副詞節)を理解する力	64%	52
関係代名詞・関係副詞を含む文を理解する力	65% @ 666666666	53
名詞を後ろから修飾・説明する構造を理解する能力	62%	54
代名詞が指すものを理解する力	45%	52
and, or で結ばれて長くなる表現を理解する力	59%	52

状況別Can Doレベル

中学英文	中学レベルの教材を読み上げた録音を聞いて
高校英文	高校レベルの教材を読み上げた録音を聞いて
日本人の授 業	日本人の先生が授業中に簡単な英語で指示するのを聞いて
ネイティブ授 業	ネイティブスピーカーの先生の英語の授業を受けて
	海外旅行中に食事・買い物のために店に入り、あなたの
海外の店員	
	言ったことに店員さんが答えるのを聞いて
歌	テンポのゆっくりした英語の歌(バラードなど)を聞いて
海外アナウ	海外旅行中に空港や駅などで出発便やプラット英語のア
ンス	ナウンスを聞いて
ニュース	海外で作成されたニュース番組を見て
映画	英語の映画を字幕なしで見て
ネイティブ会	ネイティブスピーカー同士が自然に話しているのを聞い
話	て