



NATIONAL LANGUAGE SERVICE CORPS

Language for the good of all.

Evaluating the Validity of NLSC Self-Assessment Scores

Charles W. Stansfield

Jing Gao

Bill Rivers

Overview

- **Introduction**
 - ↗ Background of NLSC Certification and Screening
 - ↗ Purpose
- **Research Design**
 - ↗ Data Sources
 - ↗ Characteristics of the Sample
 - ↗ Results of the Self-assessments and OPIs
 - ↗ Predictive Validity Study
- **Results**
- **Conclusion and Discussion**



Introduction: Background of NLSC Certification and Screening



Applicant Login

Home About Us Membership FAQ Press Room Apply Agencies Contact Us

Language for the good of all.®

NLSC Blog

On endangered and lost languages

WEDNESDAY, SEP 29, 2010 09:51

The NLSC accepts applications from any US citizen who understands English and another language ... any language. As an organization that supports Federal government agencies, we work to support as many languages as possible. But our focus is on less-common languages. And since there are between 6,000 and 7,000 languages in the world, we could list hundreds that most people have never heard of.



NLSC Member Harry interprets for the US Army in Indonesia

Use your language skills to help others!

Learn About Us

Vietnamese Ad

Our English-Language Ad as seen on CNN and Fox News Channel

Follow us on...



Introduction: Background of NLSC Certification and Screening

- NLSC is being established as a new organization to provide and maintain a standing civilian corps of certified bilinguals who will be available for service to federal government agencies as they are needed, and to state and local government agencies in time of emergency.
- Intent: fill the gap between full-time language services professionals and individuals who wish to volunteer for temporary services for short or medium term assignments.
- NLSC now has 1516 charter members
- NLSC is actively seeking speakers of 12 languages



Introduction: Background of NLSC Certification and Screening

- The NLSC must qualify applicants as part of its enrollment process.
- The NLSC uses the Federal Interagency Language Roundtable Proficiency Guidelines (the ILR scale) in speaking, reading, and listening as a basis for determining eligibility for Charter membership.
- The NLSC requirement for a qualified candidate is 3/3/3 proficiency (speaking/reading/listening).
- All NLSC applicants are screened for foreign language proficiency by asking them to complete a series of self-assessments as part of the application process. These self-assessments provide an indication of where applicants fall on the ILR scale.
- Formal assessment of English language skills is waived for applicants who attended and graduated from an accredited high school or college in the U.S. for at least three years.



Introduction: Background of NLSC Certification and Screening

- In the current pilot program, all applicants fill out a basic application form, respond to a language-background questionnaire and complete a two-part self-assessment form.
- Can-do statements: commonly referred to as Can-do scales in the language testing literature.
- Global assessment: simplified set of ILR skill level descriptions. The candidate will read the description for each skill and select the one that best describes his or her language proficiency in that skill.
- If the candidate demonstrates proficiency at ILR level 3 or higher on the predicted language proficiency rating, he or she will undergo formal testing of language skills.



Purpose of the study

- gather evidence to support the valid interpretation of two types of self-assessment instruments used in screening applicants at NLSC.
- contribute to the usefulness, acceptance, and sustainability of these assessments.
- four questions of potential concern to the NLSC administrators and applicants are posed and relevant findings are reported under each question.



Research Designs

- Data Sources
 - ä four skills are assessed: listening, speaking, reading and writing (The data for the writing subset of Can-do statements are not available).
 - ä The 158 Can-do statements (DD Form 2933, Version 4, Sep 2009) describe concrete tasks: 40 listening, 48 speaking, 32 reading, and 38 writing.
 - ä Global assessment: the plus level is interpreted as 0.6 higher than the baseline level.

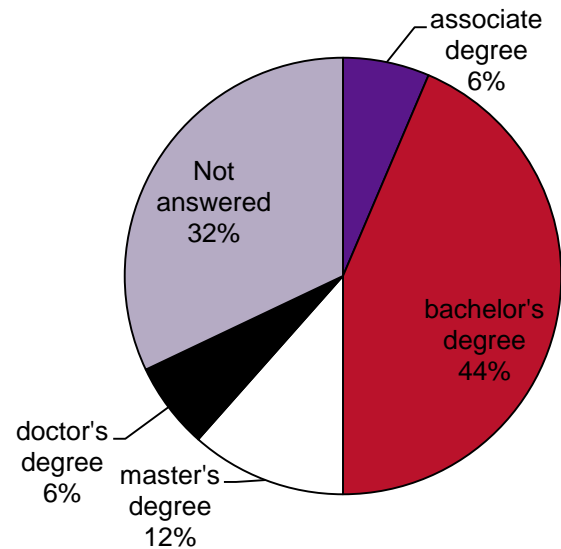
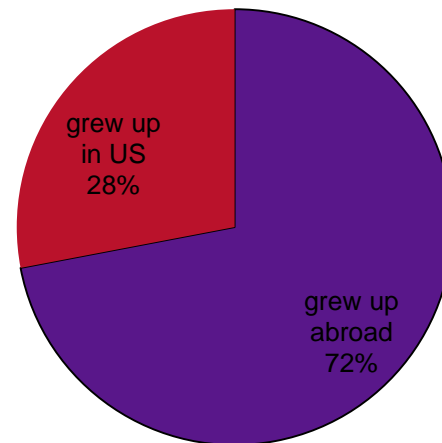
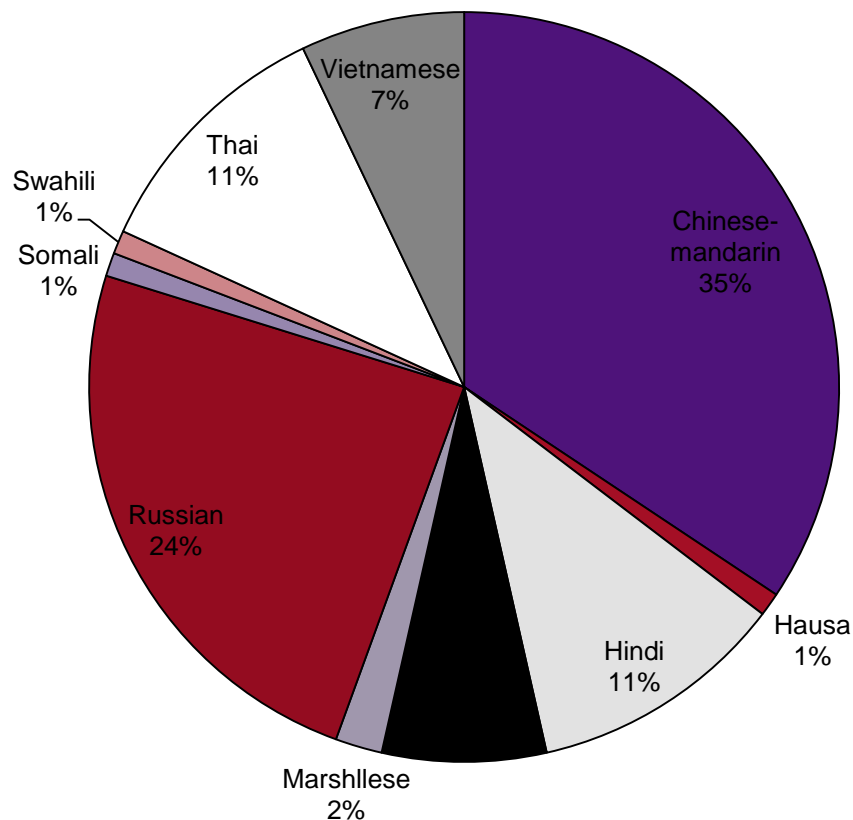


Characteristics of the Sample

- **Background questionnaire:**
- General information (age, name, address of applicants)
- Language Experience (target language, native language, and where they learned the language)
- General Information (citizenship, willingness to undergo a background investigation, etc)
- Education Information (high school, college, and other qualifications)
- Applicant certification



Characteristics of the Sample



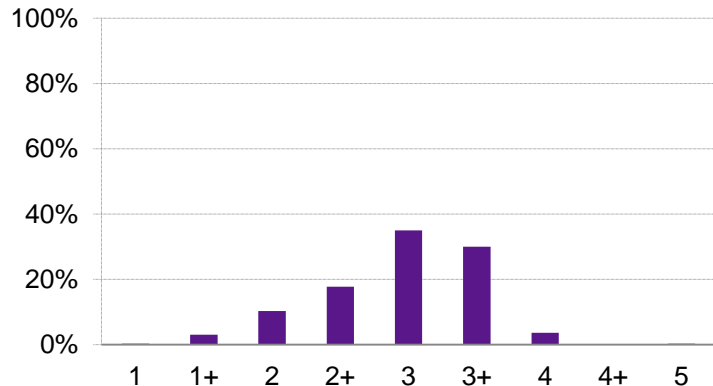
Characteristics of the Sample

Table 1: Score Distributions of Self-assessments and the OPI (among 249 candidates)

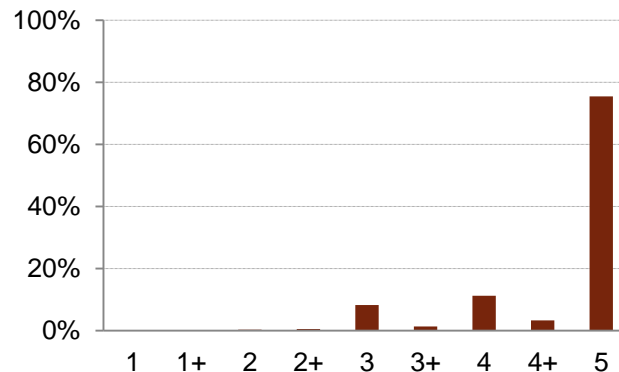
ILR Level	1	1+	2	2+	3	3+	4	4+	5
OPI									
N	1	11	36	62	122	103	13	0	1
%	0.30%	3.1%	10.3%	17.8%	35.0%	30%	3.7%		0.30%
Can-do: listening									
N	0	0	1	2	29	5	36	11	263
%			0.3%	0.6%	8.3%	1.4%	11.3%	3.3%	75.4%
Can-do: speaking									
N	0	0	7	8	51	10	50	6	215
%			2.0%	2.3%	14.6%	2.9%	14.3%	1.7%	61.6%
Can-do: reading									
N	0	1	5	7	54	8	31	3	238
%		0.30%	1.4%	2%	15.5%	2.3%	8.9%	0.9%	68.2%
Global: listening									
N	0	0	0	3	21	31	30	35	225
%				0.90%	5.90%	9.30%	7.70%	9.60%	65.30%
Global: writing									
N	4	3	10	23	33	31	21	38	171
%	1.10%	0.9%	3%	6.90%	9.9%	9.3%	6.3%	11.4%	51.2%
Global: reading									
N	0	1	4	11	27	24	29	34	216
%		0.3%	1.2%	3.2%	7.8%	6.9%	8.4%	9.8%	62.4%
Global: speaking									
N	0	0	1	11	21	27	32	37	217
%			0.3%	3.2%	6.0%	7.7%	9.2%	10.6%	62.2%

Characteristics of the Sample

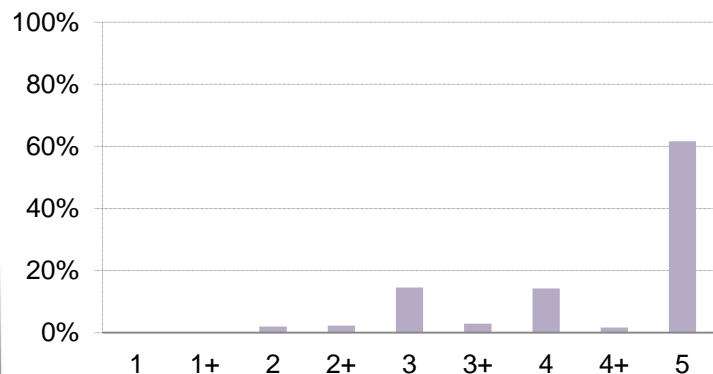
OPI Scores



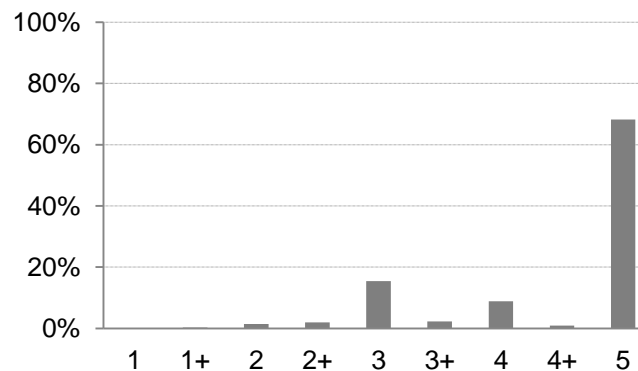
Can-do Listening



Can-do Speaking



Can-do Reading



Research Design: Predictive Validity Study

- Predictive validity: the extent to which a score on a scale or test predicts scores on some other measure, i.e., the criterion.
- For NLSC self-assessments to have predictive validity, the correlation between the self-assessment scores and formal language proficiency tests needs to be statistically significant and of at least moderate effect size.
- Oral Proficiency Interview (a carefully structured conversation between a certified interviewer and the candidate) score (OPI score) served as the criterion measure for evaluating the validity of the Self-assessments.



Research Questions

- **Research Question 1:** Among Can-do statements and global self-assessments, which generated higher self-ratings and which generated lower self-ratings?
- **Research Question 2:** Are there statistically significant correlations between self-assessment scores and the direct measures of language proficiency? What is the relationship among scores on the two types of self-assessment instruments?
- **Research Question 3:** What is the effect size and practical utility of the correlations? How do the correlations compare with those found in predictive validity studies of high stakes tests such as the GRE and the SAT?
- **Research Question 4:** What is the predictive validity of the Global self-assessments and the Can-do statements respectively in predicting an OPI score?



Research Question 1: Can-do Statements vs. Global Self-assessments

Table 3: Paired sample t-tests

		Mean	Std. Error Mean	t	df	Sig. (2-tailed)
Pair 1	CAN-DO: LISTENING-GLOBAL: LISTENING	0.065	0.03	2.47	343	0.014
Pair 2	CAN-DO: READING-GLOBAL: READING	-0.044	0.03	-1.48	344	0.139
Pair 3	CAN-DO: SPEAKING-GLOBAL: SPEAKING	-0.16	0.03	-5.102	344	0.00

- Motivation: reduce the number of forms a candidate needs to fill

statistical significance (listening & speaking) \neq practical utility

- Conclusion: The NLSC could state that self-assessment scores are generally comparable across the self-assessment instruments that assess listening and reading skills.

Research Question 2: Are there statistically significant correlations between self-assessment scores and the direct measures of language proficiency? What is the relationship among scores on the two types of self-assessment instruments?

		OPI	Can-do: Listening	Can-do: Speaking	Can-do: Reading	Global: Listening	Global: Writing	Global: Reading	Global: Speaking
OPI	Pearson Correlation	1							
	Sig. (2-tailed)								
	N	349							
Can-do: Listening	Pearson Correlation	0.426**	1						
	Sig. (2-tailed)	0.000							
	N	347	372						
Can-do: Speaking	Pearson Correlation	0.395**	0.728**	1					
	Sig. (2-tailed)	0.000	0.000						
	N	347	372	372					
Can-do: Reading	Pearson Correlation	0.454**	0.690**	0.806**	1				
	Sig. (2-tailed)	0.000	0.000	0.000					
	N	347	372	372	372				
Global: Listening	Pearson Correlation	0.535**	0.722**	0.707**	0.740**	1			
	Sig. (2-tailed)	0.000	0.000	0.000	0.000				
	N	345	369	369	369	370			
Global: Writing	Pearson Correlation	0.440**	0.602**	0.704**	0.794**	0.723**	1		
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000			
	N	334	356	356	356	356	357		
Global: Reading	Pearson Correlation	0.495**	0.679**	0.679**	0.791**	0.801**	0.850**	1	
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000		
	N	346	370	370	370	370	357	371	
Global: Speaking	Pearson Correlation	0.480**	0.698**	0.747**	0.714**	0.887**	0.747**	0.779**	1
	Sig. (2-tailed)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	346	370	370	370	370	357	371	371

** Correlation is significant at 0.01 level (2-tailed).

Research Question 3: What is the effect size and practical utility of the correlations? How do the correlations compare with those found in predictive validity studies of high stakes tests such as the GRE and the SAT?

- By convention, correlation coefficients of 0.10, 0.30, and 0.50 are termed small, moderate, and large respectively in terms of their effect size (Cohen, 1988).
- the correlation coefficient results are not corrected for the restriction of range
- Heilenman (1990): $r=.33$
- Ross (1998) : meta-analaysis $r=0.61$



Research Question 4: What is the predictive validity of the Global self-assessments and the Can-do statements respectively in predicting an OPI score?

Modeling process:

1. Have all can-do assessment scores and global assessment scores as IV (independent variables), and OPI as DV (dependent variable). –Not working well

		MODEL 1	MODEL 2
PREDICTED VARIABLE	OPI	X	X
	CAN-DO: LISTENING	X	
	CAN-DO: SPEAKING	X	
	CAN-DO: READING	X	
PREDICTOR VARIABLES	GLOBAL: LISTENING		X
	GLOBAL: WRITING		X
	GLOBAL: READING		X
	GLOBAL: SPEAKING		X

2. Two models

Research Question 4: What is the predictive validity of the Global self-assessments and the Can-do statements respectively in predicting an OPI score?

		Coefficients(a)				
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.11	0.21		5.38	0.00
	CAN-DO: LISTENING	0.21	0.07	0.23	3.16	0.00
	CAN-DO: SPEAKING	-0.02	0.06	-0.03	-0.31	0.76
	CAN-DO: READING	0.22	0.06	0.32	3.88	0.00

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	0.69	0.20		3.41	0.00
	GLOBAL: LISTENING	0.38	0.10	0.41	3.83	0.00
	GLOBAL: WRITING	0.00	0.05	0.00	-0.04	0.97
	GLOBAL: READING	0.16	0.08	0.21	2.03	0.04
	GLOBAL: SPEAKING	-0.03	0.09	-0.04	-0.37	0.71

R-Square: .231

Using GMAT to predict first year GPA:
R-Square: .213

R-Square: .305



Discussion and Conclusion

- Overall, the implications of this study are that the Can-do statements and the global self-assessments were valid instruments for the measurement of language skills in target languages and should remain as part of the NLSC screening process.
- Samples : candidates already admitted to NLSC membership, they underestimate the true correlations that would be obtained if all candidates for whom self-assessment data were available.
- We suggest collecting the data on the non-admitted candidates and correcting the correlations due to the restriction of range in the self-assessments.



References

- Heilenman, L. K. (1990). Field test for the ITC guidelines for adapting educational and psychological tests. *European Journal of Psychological Assessment*, 15 (3), 270-276.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15, 1-20.
- Sireci, S. & Miller, E. (2006). Evaluating the Predictive Validity of Graduate Management Admission Test scores. *Educational and Psychological Measurement*, 66 (2), 305-317

