

C - Test و سنجش توانایی عمومی زبانی
اثر: دکتر علی افخمی
دانشیار دانشکده ادبیات و علوم انسانی دانشگاه تهران
و فردوس جمالی
(از ص ۵۰۹ تا ۵۲۳)

چکیده:

در تحقیق حاضر، گونه‌های مختلف C - Test بعنوان آزمون‌هایی برای سنجش توانایی عمومی زبانی مورد بررسی قرار گرفتند. در آغاز C - Test بعنوان گونه اصلاح شده آزمون بسته (cloze) به رشته آزمون سازی زبان وارد شد و جایگاه خود را پیدا کرد. اولین C - Test از طریق حذف نیمه دوم کلمات (بصورت یک کلمه در میان) در یک متن چند پاراگرافی ساخته شد، عمل حذف از جمله دوم متن آغاز می‌شد. بررسی‌های متعدد نشان دادند که این آزمون می‌تواند توانایی عمومی زبانی را همانند تافل و آزمون بسته اندازه بگیرد. هدف تحقیق حاضر این بود که یکبار دیگر روی نتایج تحقیقات قبلی صحه بگذارد. بدین منظور، ده گونه مختلف C - Test با استفاده از روشهای حذف گوناگون ساخته شدند که یکی از این ده نوع بر اساس همان قاعده اول ساخته شد. این ده نوع C - Test به‌مراه یک تافل و آزمون بسته استاندارد در میان حدود ۵۰۰ نفر دانشجوی سال آخر رشته زبان انگلیسی اجرا شد. نتایج تجزیه و تحلیلهای آماری نشان داد که C - Test می‌تواند تواناییهای متعددی را اندازه بگیرد که یکی از آنها احتمالاً توانایی عمومی زبانی است و در این راستا، نمی‌توانند این توانایی را بطور یکسان اندازه بگیرند.

واژه‌های کلیدی: C - Test، سنجش، توانایی عمومی زبانی، ارزش علمی.

مقدمه:

اسپالسکی (spolsky) (۱۹۷۸) مکاتب عمده در آزمون سازی زبان در قرن بیستم را به سه دسته تقسیم می‌کند: غیر علمی (pre-scientific)، ساخت گرا - روان شناختی (psychometric - structuralist)، و زیانشناسی اجتماعی - تلفیقی (integrative - sociolinguistic).

آزمون‌هایی که نماینده گروه اول هستند در کلاسهای درس بسیار متداولند و معلم تنها شخصی است که در مورد طرح و اجرای این آزمونها تصمیم می‌گیرد. دومین مکتب اساس ساخت آزمونهای علمی و استاندارد است، بارزترین نمونه این آزمونها تست تافل است. از نقطه نظر مکتب سوم، استفاده از زبان یک فرایند کلی (holistic) و تلفیقی است و بهمین دلیل آزمونهای این دوره هم تلفیقی می‌باشند. آزمونسازی زبان در مرکز آموزش زبان قرار دارد و نمی‌تواند جدای از آن تعریف شود.

آزمون سازی زبان نه تنها اهداف نهایی در آموزش زبان را بیان میکند بلکه موفقیت معلمان و فراگیران را در دستیابی به آن اهداف می‌سنجد. بعلاوه، آزمون سازی زمینه را برای اجرای آزمایش و تحقیق در آموزش زبان و یادگیری واکنساب زبان فراهم می‌کند. پس بجاست که آزمون سازی را بعنوان یکی از شاخه‌های زیانشناسی کاربردی به حساب آوریم. دیویس (Davis) (۱۹۹۰) از آزمون سازی بعنوان لبه زیانشناسی کاربردی نام می‌برد، بعقیده وی زیانشناسی کاربردی بدون آزمون سازی معنا و مفهومی ندارد.

یکی از پیشرفتهای آزمون سازی در مکتب سوم استفاده از اصل «کاهش حشو» (reduced redundancy) است.

منظور از این اصل این است که بزرگسالان با استفاده از دانش خود از قوانین، الگوها و اصطلاحات زبانی، قادرند جاهای خالی را در زبان گفتار یا نوشتار پرکنند.

بنابر این فراگیران یک زبان که هنوز توانش زبانی گسترده‌ای ندارند در بازسازی پیامهای ناقص دچار مشکل می‌شوند. اصل کاهش حشو توسط اسپالسکی (۱۹۶۸) و آلر (oller) (۱۹۷۶) ارائه شد.

آزمون بسته برای اولین بار در سال ۱۹۵۳ توسط تیلور (Taylor) بعنوان ابزاری براس سنجش درجه خوانایی (readability) متون مختلف ارائه شد ولی اندکی بعد برای اندازه‌گیری توانایی خواندن و درک مفاهیم در زبان اول و زبان دوم بکار رفت. مطالعات متعدد نشان داد که آزمون بسته دارای اعتبار (reliability) و ارزش (validity) علمی بالایی است و می‌تواند برای سنجش توانایی عمومی زبان مورد استفاده قرار گیرد. با اینحال تعدادی از تحقیقات نشان داده‌اند که این آزمون دارای نقایصی است. آلدرسون (Alderson) (۱۹۸۳)، فرهادی (۱۹۶۶ و ۱۹۸۳)، وکلاین - برایلی و راتز (Klein - Braley & Raatz) (۱۹۸۴) ارزش علمی آزمون بسته را زیر سوال می‌برند. آلدرسون (۱۹۸۷) نشان داد که تغییر سطح دشواری متن یا تغییر فواصل حذف بر نتایج حاصله تاثیر می‌گذارد. این قبیل انتقادات کلاین - برایلی و راتز (۱۹۸۱) را بر آن داشت که C - Test را بعنوان مدل ارتقا یافته آزمون بسته ارائه کنند. C - Test اولیه با حذف نیمه دوم کلمات بصورت یک کلمه در میان و با شروع از کلمه دوم در جمله دوم متن ساخته می‌شد.

۲- مروری بر تحقیقات انجام شده در مورد C - Test :

بیشتر تحقیقاتی که روی C - Test صورت گرفته حاکی از این است که این آزمون از اعتبار و ارزش علمی بالایی برخوردار است. بعنوان مثال کلاین - برایلی و راتز (۱۹۸۵، ۱۹۸۴، ۱۹۸۱) معتقدند که C - Test ابزار مناسبی برای سنجش توانایی عمومی زبانی است.

آنها فواید استفاده از C - Test را بدین ترتیب ذکر می‌کنند :

- (۱) ساختن و نمره‌گذاری C - Test آسان است .
 - (۲) سخن‌گویان بومی زبان نمره کامل را در C - Test کسب می‌کنند .
 - (۳) بخشهای حذف شده نماینده کل متن می‌باشند .
 - (۴) حتی مطالبی که قبلاً مورد آزمایش قرار نگرفته‌اند اعتبار و ارزش بالایی بدست می‌دهند .
 - (۵) C - Test دارای ارزش صوری (face validity) است .
- راتز (1985 c) C - Test را در کلاسهای درس انگلیسی و آلمانی (بعنوان زبان مادری) در سطوح مختلف اجرا کرد میزان ارزش و اعتبار علمی آنها را اندازه گرفت. نتیجه این شد که C - Test برای کلیه سطوح آزمون معتبری است. راتز (1985 d) ارزش فاکتوری (factorial validity) C - Test را برای آلمانی زبانان اندازه گرفت. او بین نمرات C - Test و نمرات کلاسی بچه‌ها ضریب همبستگی بالایی یافت، در مورد دانش‌آموزان انگلیسی و اسپانیولی هم نتیجه مشابهی بدست آمد. همچنین نمرات C - Test ضریب همبستگی بالایی با تستهای هوش و تمرکز بدست دادند. گرات جان (1986 a,b , 1987 b,c) و گرات جان و استمر (Grotjahn & Stemmer) (۱۹۸۴، ۱۹۸۵) برای زبان فرانسه C - Test طراحی کردند. این تست دارای اعتبار بالایی بود ولی در مورد ارزش علمی آن ارزشیابی دقیقی صورت نگرفت. اما از سوی دیگر وارد (Ward) (۱۹۸۷) دریافت که C - Test در زبان فرانسوی هم دارای اعتبار و ارزش علمی بالایی است و از قدرت تمیز (discrimination) زیادی برخوردار است. جعفر پور (۱۹۹۵) هم به این نتیجه رسید که C - Test از بسیاری جهات بر آزمون بسته برتری دارد. او بیست‌گونه از C - Test را با روشهای مختلف ساخت، نتایج تحلیلهای او بدین قرار است :
- (۱) ساختن و نمره‌گذاری C - Test آسان است، ولی
 - (۲) سخن‌گویان بومی زبان همواره نمره کامل را کسب نمی‌کنند.

(۳) بخشهای حذف شده نماینده کل متن نیستند، چون شیوه‌های مختلف حذف باعث می‌شود C - Test های مختلفی بدست آید.

(۴) مطالبی که قبلاً مورد آزمایش قرار نگرفته‌اند اعتبار علمی بالایی را نشان می‌دهند اما در مقایسه با آزمون بسته دارای ارزش عملی چندانی نیستند.

(۵) C - Test دارای ارزش صوری نیست.

اگر تحقیقات صورت گرفته روی C - Test نشان می‌دهند که این آزمون توانایی عمومی زبانی را اندازه‌می‌گیرد. از جمله کوهن، سیگال و ویس (Cohen , Segal & Weiss) (۱۹۸۵) یک C - Test به زبان عبری ساختند و اعتبار علمی، قدرت تمیز، ضریب همبستگی با سایر آزمونهای توانایی زبانی، ضریب همبستگی با ارزشیابیهای معلمان، و تفاوت آن را با آزمون بسته از نقطه نظر فرایند پاسخ‌گویی بررسی کردند. نتایج تحلیل داده‌ها نشان دادند که C - Test دارای اعتبار و ارزش علمی بالایی بوده و بخوبی بین دانش‌آموزان سطوح مختلف تمیز می‌دهد. به همین دلیل این محققان ابراز کردند که این آزمون ابزار مناسبی برای سنجش توانایی عمومی زبانی است.

لین (Lin) (۱۹۸۷) دریافت که بین C - Test و دیکته ضریب همبستگی بالایی وجود دارد.

ژانگ (Zhang) (۱۹۸۷) C - Test را برای سنجش توانایی عمومی زبانی در زبان چینی بکار برد.

نگیشی (Negishi) (۱۹۸۷) دریافت که C - Test با بخش گرامر، واژگان و خواندن و درک مفاهیم تست MELAB (Michigan English Language Battery) (نوعی تست استاندارد شبیه تافل) ضریب همبستگی بالایی دارد.

دورینه و کاتونا (Dorney & Katona) (۱۹۹۲) هم C - Test را برای مجاری‌هایی که انگلیسی را بعنوان زبان خارجی می‌آموختند بکار بردند. آنها هم به این نتیجه

رسیدند که C - Test ابزاری مناسب و انعطاف پذیر برای سنجش توانایی زبانی در سطوح مختلف است. تحقیقات ذکر شده در بالا و سایر تحقیقاتی از این دست، همگی با کلاین - برابلی (۱۹۹۶ و ۱۹۸۵ و ۱۹۸۴) در این موضوع که C - Test توانایی عمومی زبانی را می‌سنجد، همراهی می‌کنند. هدف از تحقیق کنونی این بود که بفهمیم آیا در مورد فارسی زبانان هم C - Test می‌تواند بعنوان ابزاری برای سنجش توانایی عمومی زبانی بکار رود یا نه .

۳- روش کار :

۳-۱ نمونه ۴

افراد شرکت کننده در این تحقیق عبارت بودند از حدود ۵۰۰ نفر دانشجوی سال آخر رشته زبان انگلیسی که در دانشگاههای تهران مشغول به تحصیل بودند .

۳-۲ ابزار

ابزار اجرای این تحقیق عبارت بود از سه نوع تست : تست استاندارد تافل ۱۹۹۵، آزمون بسته استاندارد و ده نوع C - Test . تست تافل و آزمون بسته برای سنجش ارزش علمی C - Test ها متنی با درجه دشواری حدود ۱۹ از دایرةالمعارف بریتانیکا انتخاب شد. آزمون بسته استاندارد با حذف هر هفت کلمه در میان از همین متن ساخته شد که جمعاً ۳۱ کلمه حذف شدند. C - Test ها هم از همین متن و به روش زیر ساخته شدند. پنج نوع اول C - Test ها با حذف نیمه دوم (SH) (second Half) کلمات و بصورت یک کلمه در میان، دو کلمه در میان، سه کلمه در میان، چهار کلمه در میان و پنج کلمه در میان ساخته شدند و با کدهای SH6,SH5,SH4,SH3,SH2 نامگذاری شدند. پنج نوع دوم با حذف نیمه اول (FH) (first Half) کلمات و به روش بالا ساخته شدند و با کدهای FH6,FH5,FH4,FH3,FH2 شناسایی شدند. لازم بتوضیح

است که تعداد حذف در پنج نوع اول و هم پنج نوع دوم بترتیب ۱۰۰، ۶۸، ۵۱، ۴۱، ۳۴ می باشد.

۳-۳ اجرای آزمونها

تستهای آماده شده ابتدا در یک گروه سی نفری از دانشجویان سال آخر رشته زبان انگلیسی دانشگاه علامه طباطبایی مورد پیش آزمون قرار گرفتند. در مرحله اصلی تحقیق، ابتدا تست تافل و آزمون بسته در میان همه افراد نمونه اجرا شد و پس از دو هفته Test - C ها هم اجرا شدند. جدول ۱ چگونگی اجرای آزمونها را بروشنی نشان می دهد:

جدول ۱

گروه	تعداد	نوع تست
1	47	TOFFEL + CLOZE + SH2
2	49	TOFFEL + CLOZE + SH3
3	51	TOFFEL + CLOZE + SH4
4	48	TOFFEL + CLOZE + SH5
5	49	TOFFEL + CLOZE + SH6
6	46	TOFFEL + CLOZE + FH2
7	51	TOFFEL + CLOZE + FH3
8	46	TOFFEL + CLOZE + FH4
9	46	TOFFEL + CLOZE + FH5
10	48	TOFFEL + CLOZE + FH6

۳-۴ تحلیل داده‌ها

تجزیه و تحلیل داده‌ها طی مراحل زیر صورت گرفت: ابتدا آمار توصیفی مقدماتی در مورد همه ابزارها صورت گرفت. در مورد آزمون بسته و تافل، میانگین نمرات و میزان انحراف معیار در همه گروهها تقریباً یکسان بود، لذا از ذکر آنها در جدول خودداری می‌شود. سپس تحلیل واریانس یکطرفه بین همه نمرات آزمون بسته، همه نمرات تافل، و همه نمرات C - Test انجام شد. در مورد تافل و آزمون بسته مقدار F مشاهده شده از F معیار کمتر بود که این نشان می‌دهد تفاوت معناداری بین گروهها از نقطه نظر عملکرد روی تافل و آزمون بسته وجود ندارد. اما در مورد C - Test ها نتیجه دیگری بدست آمد که در جدول مربوطه شرح داده خواهد شد. اعتبار علمی همه آزمونها از طریق فرمول KR-21 بدست آمد. ضریب همبستگی هر کدام از C - Test ها با تافل و آزمون بسته و همچنین ضریب همبستگی بین تافل و آزمون بسته محاسبه شد. سرانجام تحلیل فاکتور بین همه C - Test ها، تافل و آزمون بسته صورت گرفت. نتایج این تحلیلها در جداول زیر آمده است:

جدول ۲: آمار توصیفی C - Test ها

	SH2	SH3	SH4	SH5	SH6	FI2	FI3	FI4	FI5	FI6
X	42.6	41.84	34.94	25.17	19.96	54.22	42.71	34.04	24.72	19.13
SD	15.51	6.58	4.4	3.96	5.26	11.92	8.42	7.14	6.1	5.17
Min.	28	26	25	16	10	36	24	14	16	15
Max.	71	53	43	31	32	81	60	45	35	28

طبق جدول فوق، ظاهراً بهترین عملکرد روی SH4 بوده است.

جدول ۳: تحلیل واریانس یکطرفه برای C - Test ها

Source	d.f	SS	MS	F RATIO	Fprob
between	9	23287.73	2587.53	16.18	.000
within	471	75143.43	159.88		
total	480	98431.16			

در $P=0.05$ و $d.f(9,471)$ ، مقدار F معیار برابر با ۱.۹۰ است. طبق جدول (۳) مقدار F مشاهده شده از F معیار بسیار بزرگتر است و این بدان معناست که گروهها عملکردهای مختلفی روی C - Test داشته‌اند. بعبارت دیگر تفاوت معناداری بین گروهها از نقطه نظر عملکرد روی C - Test آنها داشته‌اند. بعبارت دیگر تفاوت معناداری بین گروهها از نقطه نظر عملکرد روی C - Test آنها وجود دارد.

جدول ۴: اعتبار علمی همه آزمونها

Test	1	2	3	4	5	6	7	8	9	10
cloze	.84	.25	.15	.22	.46	.21	.46	.55	.36	.91
TOEFL	.84	.86	.44	.68	.86	.65	.87	.87	.80	.81
C-TEST	.85	.75	.70	.75	.91	.83	.86	.86	.90	.90

جدول ۵: ضریب همبستگی C - Test ها با تافل و cloze

Variable	SH2	SH3	SH4	SH5	SH6	FH2	FH3	FH4	FH5	FH6
TOEFL	.46	.77	.36	.56	.66	.43	.80	.61	.35	.52
CLOZE	.37	.53	.64	.54	.91	.63	.44	.72	.63	.40

طبق جدول (۴) C - Test ها و تافل از اعتبار علمی بالایی برخوردارند. جدول (۵) هم نشان می‌دهد که بالاترین ضریب همبستگی از یک سو بین SH6 و CLOZE و از سوی دیگر بین FH3 و تافل است.

نتایج تحلیل فاکتور نشان داد که آزمون بسته و بخشهای مختلف تست تافل (شامل گرامر، واژگان و خواندن و درک مفاهیم) همگی زیر یک فاکتور جمع

می شوند. این نتیجه غیر منتظره نبود چون عقیده عمومی بر این است که هر دوی این آزمونها توانایی عمومی زبانی را اندازه می گیرند. از میان ده نوع C - Test، ۶ تای آنها زیر یک فاکتور قرار گرفتند و این فاکتور متفاوت از فاکتوری بود که تافل و آزمون بسته زیر آن قرار گرفته بودند. چهارتای دیگر C - Test ها (SH6, SH2, FH3, FH2) رفتار متفاوتی از سایرین نشان دادند. چون دلیل فاعل کننده ای مبنی بر علت تفاوت رفتار بعضی C - Test ها از سایرین، در مطالعات و تحقیقات گذشته روی C - Test یافت نشد تصمیم بر این شد که این چهار C - Test از محاسبات حذف شده و عملیات تحلیل فاکتور مجدداً تکرار شود.

جدول (۶) تحلیل فاکتور تعدیل شده را نمایش می دهد:

جدول ۶: تحلیل فاکتور تعدیل شده

Variable	Comp	FH4	FH5	FH6	SH3	SH4	SH5	View	Gram	Read
Factor1	.67	*	*	*	*	*	*	.77	.67	.78
Factor2	*	.69	.71	.58	.71	.73	.68	*	*	*

حال اگر نتایج جدول ۶ را با جداول ۴ و ۵ مقایسه کنیم می بینیم که در میان شش C - Test باقیمانده، FH4 دارای اعتبار و ارزش علمی بالاتری است و گونه مناسبتری برای سنجش توانایی عمومی زبانی به شمار می رود.

نتیجه:

همانگونه که پیشتر ذکر شد، تحقیقات متعدد نشان داده اند که C - Test ابزاری با اعتبار و ارزش علمی بالاست که قابلیت سنجش توانایی عمومی زبانی را داراست. اما نتایج تحقیق کنونی یافته های پیشین را تایید نمی کند. درست است که اعتبار

علمی C - Test ها بالا بود (و این امر با توجه به همسان بودن گروهها قابل پیش بینی بود)، ولی C - Test ها از ارزش علمی چندان بالایی برخوردار نبودند. نتیجه دیگری که از تحقیق حاضر بدست می آید این است که همه گونه های C - Test قابلیت سنجش توانایی عمومی زبانی را ندارند و در واقع C - Test ها لزوماً و بطور اخص این توانایی را اندازه نمی گیرند. آنها ممکن است قابلیت های متعدد و مختلفی را اندازه بگیرند و یکی از اینها احتمالاً توانایی عمومی زبانی است .

References

- 1- Alderson.J.C. (1983). The cloze procedure and proficiency in English as a foreign language. In Oller. J.W.Jr (Ed). Issues in language testing research. Rowley. Mass.: Newbury House. 205-217.
- 2- Caldicott, G.A. (1987). The C-test and text difficulty. Language Testing Update. 4.24.
- 3- Cleary, C. (1988). The C-test in English: left-hand deletions. RELC Journal. 19.26-37.
- 4- Cohen, A.D. Segal.M. & Bar-Siman-Tov. R.W.(1985). The C-test in Hebrew. Language Testing. 1. 121-127.
- 5- Davues.A.(1990). Principles of language testing. Basil Blackwell Ltd.
- 6- Farhady, H. (1983): New directions for ESL proficiency testing.

- In Oller 1983b: 253-268.
- 7- Grotjahn, R. (1986b). Test validation and cognitive psychology: some methodological considerations. *Language Testing*, 3, 159-185.
- 8- Grotjahn, R. (1987). Ist der C-test Lesetest. In Addison & Vogel (Ed.) *Lehren und lernen von Fremdsprachen in Studium*. Bochum: AKS-Verlog: 230-248.
- 9- Grotjahn, R., & Stemmer, B. (1985). On the development and evaluation of a C-test for French. *Fremdsprachen und Hochschule*, 13/14, 101-120.
- 10- Hatch, E., & Farhady, H. (1982). *Research design and statistics for applied linguistics*. Rahanma Publications.
- 11- Jafarpur, A. (1995). Is C-testing superior to cloze? *Language Testing*, 12, 194-216.
- 12- Kamimoto, T. (1992). An inquiry into what a C-test measures. *Fukuoka Women's Junior College Studies*, 30(12), 67-79.
- 13- Kamimoto, T. (1993). Tailoring the test to fit the students: improvement of the C-test through classical item analysis. *Fukuoka Women's Junior College Studies*, 30(11), 47-61.
- 14- Katona, L., & Dornyei, Z. (1992). Validation of the C-test amongst Hungarian EFL learners. *Language Testing*, 9, 187-206.
- 15- Katona, L., & Dornyei, Z. (1993). The C-test; A teacher-friendly way to test language proficiency. *English*

Teaching Forum. 31(1). 34-36.

16- Klein-Braley, C. (1981). Empirical validation of cloze tests: an examination of the validity of cloze tests as tests of general language proficiency in English for German university students. Unpublished Ph.D. dissertation. Duisburg: Universitat Duisburg.

17- Klein-Braley, C. (1985a). A cloze-up on the C-test: a study of the construct validation of authentic tests. *Language Testing*, 2(1), 76-104

18- Klein-Braley, C. (1985c). C-tests and construct validity. In Klein-Braley and Raatz (Eds.), *Lehren und Lernen von Fremdsprachen in studium Bochum*: AKS-Verlag, 230-248.

19- Klein-Braley, C. (1984). Language testing with the C-test: a linguistic and statistical investigation into the strategies used by C-test takers. and the prediction of C-test difficulty. Unpublished Habilitationsschrift University Duisburg.

20- Klein-Braley, C. (1985). Appraisal of the C-test in the context of cloze family. In Wolfgang B'rner & Klaus Vogel (Eds.), *Der text in Fremdsprachen anterricht*, (197-213). Bochum: AKS.

21- Klein-Braley, C. (1987). C-tests in the context of reduced redundancy testing: an appraisal. *Language Testing*, 14(1), 47-84.

22- Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-test. *Language Testing*, 1, 134-146.

23- Klein-Braley, C., & Raatz, U. (1985). Reduced redundancy as

- an approach to language testing. Bochum:AKS.
- 24- Klein-Braley, C., & Raatz, U. (1996). Introduction to language testing and C-tests. In Coleman (Ed.) (1996), University language testing and the C-test.
- 25- Proceedings of a conference held at the university of Portsmouth in April 1995. Portsmouth: University.
- 26- Koberl, J., & Sigott, G. (1994). Adjusting C-test difficulty in German. In Grotjahn (1994) (Ed.), *Der C-test: Theoretische Grundlagen und Praktische Anwendungen*, (179-192). Bochum: Brockmeyer.
- 27- Lin,Z. (1987). Dictation and the C-test: which is better? *Language Testing Update*, 4,20.
- 28- Negishi, N. (1987). The C-test: an integrative measure? *Language Testing Update*,4,21.
- 29- Oller, J. W. (1979). *Language tests at school: a pragmatic approach*. Longman.
- 30- Oller, J. W. (1973b). *Issues in language testing research*. Rowley, Mass.: Newbury House.
- 31- Raatz, U. (1985c). Investigating dimensionality of language tests, a new solution to an old problem. In Klein-Braley & von Essen (Eds.), *Practice and problems in language testing* 8, (123-136).

- 32- Raatz, U. (1985d). Tests of reduced redundancy: the C-tests, a practical example. In Kohonen & Pitkinen (Eds.), *Language testing in schools*. Tampere: AfinaLA.
- 33- Sigott, G. (1994). Language test validity. *AAA-Arbeiten aus Anglistik und Amerikanistik*, 19(2), 287-294.
- 34- Sigott, G. (1995). The C-test: some factors of difficulty. *AAA-Arbeiten aus Anglistik und Amerikanistik*, 20(1), 43-53.
- 35- Sigott, G., & Koberl, J. (1996). Deletion pattern and C-test difficulty across languages. In Grotjahn (1996) (Ed.), *Der C-test: Theoretische Grundlagen und Praktische Anwendungen*, (159-172). Bochum: Brockmeyer.
- 36- Spolsky, B. (1978). *Advances in language testing: series 2, Approach to language testing*. VA: Center for Applied Linguistics.
- 37- Spolsky, B. (1975). *Measured words*. Oxford: Oxford University Press.
- 38- Stemmer, B. (1991). *What's on a C-test taker's mind: mental processes in C-test taking*. Bochum: Brockmeyer.
- 39- Ward, M. (1987). A study of the C-test in French. *Language Testing Update*, 4m 23.
- 40- Zhang, L.Z. (1987). The C-test in Chinese. *Language Testing Update*, 4,23.

