

تأثیر کوییزهای متناوب لغت بر فراگیری دانش لغوی دانشجویان زبان انگلیسی میناب

فخرالسادات بیدکی *

چکیده

این تحقیق، تأثیر کوییزها بر میزان فراگیری لغت دانشجویان زبان انگلیسی دانشگاه آزاد اسلامی میناب را بررسی می‌کند. جمعا ۶۰ نفر دانشجوی زن و مرد انتخاب و به ۲ گروه مساوی تقسیم شدند. همسانی ۲ گروه از طریق تحلیل آماری نمرات آزمون مهارت تعیین سطح آکسفورد ۲۰۰۷ اثبات شد. از چهار نوع ابزار برای جمع آوری داده‌ها استفاده شده است: آزمون مهارت زبان، پیش-آزمون، پس-آزمون و پس-آزمون با تأخیر. به غیر از آزمون مهارت، در ابتدا برای اثبات یکسان بودن گروه‌ها از لحاظ دانش لغوی مربوط به کتاب تدریس شده (American Reading)، پیش-آزمون از هر دو گروه گرفته شد. سپس گروه آزمایش ۸ کوییز (تست کوچک) لغت دریافت کردند. در آخر، پس-آزمون گرفته شد. تی-تست نمونه‌های مستقل بر روی نتایج پس-آزمون و پیش-آزمون انجام شد تا تأثیر کوییزهای متناوب لغت بر میزان یادگیری لغت شرکت کنندگان تعیین شود. نتایج این تحقیق مشخص کرد که کوییزهای لغت تأثیر بسزایی بر میزان یادگیری لغوی شرکت کنندگان داشته است. یک ماه بعد، از گروه آزمایش پس-آزمون با تأخیر گرفته شد. انواع مقیاس مکرر بر روی نتایج پس-آزمون با تأخیر اجرا شد تا کارایی کوییزهای متناوب در درازمدت بررسی شود. نتایج نشان داد یادگیری آن‌ها پایدار بوده است.

کلید واژه‌ها: تست، ارزشیابی زبان، واشبک، کوییز، دانش لغت

مقدمه

هدف اصلی تمامی مراکز آموزشی، پیشرفت دانش‌آموزان است و آن‌ها برای رسیدن به این هدف

همه ابزارها را امتحان می‌کنند. با توجه به نقش آزمون در یادگیری زبان، یکی از مهمترین موضوعاتی که به تحقیق بیشتری نیاز دارد، بررسی تأثیر آزمون بر میزان یادگیری زبان دانش‌آموزان است. بااینکه تحقیقات زیادی بر روی واشبک (washback) انجام شده است، تحقیقات تجربی کمی در ایران صورت گرفته که گویای وضعیت کلاسهایی باشد که در آن از کوییزهای دست ساخته معلم (که یکی از متداولترین ابزار است) استفاده می‌شود و تأثیر این کوییزها را بر یادگیری نشان دهد.

Madson (۱۹۳۴) بیان کرده که ارزشیابی، یکی از مهم‌ترین بخش‌های هر تجربه یاددهی-یادگیری است. هدف اصلی ارزشیابی از درس داده شده، فراهم کردن بازخورد است که بدون آن نه معلم و نه دانش آموز پیشرفت چندانی نمی‌کنند. تلاشهای زیادی برای ترکیب انواع ابزارها و فعالیت‌های مختلف به منظور افزایش میزان یادگیری دانش‌آموزان انجام شده است. Heaton (۱۹۸۸) معتقد است که آزمون‌ها به عنوان ابزاری برای تقویت یادگیری و ایجاد انگیزه در دانش‌آموزان و یا برای ارزیابی رفتار آن‌ها در تکالیف ساخته شده اند. هدف تحقیق حاضر، کشف تأثیر کوییزهای متناوب لغت بر میزان فراگیری دانش لغوی دانشجویان زبان انگلیسی دانشگاه آزاد میناب است.

پیشینه ادبی

چهار چوب نظری

اصطلاحاتی که برای تأثیر امتحان استفاده شده:

تاکنون اصطلاحات گوناگونی برای بیان مفهوم تأثیر امتحان به کار گرفته شده است.

Gosa (۲۰۰۴، p. ۲۵) به برخی از مشهورترین آن‌ها اشاره می‌کند:

آموزش بر گرفته از سنجش (Popham, 1987)

بک واش (Hughes, 1989)

اعتبار واشبک (Morrow, 1986; Weir, 1990)

اعتبار اصولی (Fredericksen & Collins, 1989)

تنظیم برنامه درسی (Smith, 1991)

واشبک (Alderson & Wall, 1993)

تأثیر (Wall, 1997)

اگرچه تفاوت‌هایی بین اصطلاحاتی که نویسندگان به کار برده‌اند دیده می‌شود ولی به طور کلی، همه آن‌ها به یک پدیده اساسی اشاره می‌کنند: رابطه بین آزمون، یاددهی و یادگیری. محقق تصمیم گرفته از واژه "واشبک" که Alderson & Wall (۱۹۹۳) به کار برده‌اند و توسط متخصصان نیز توصیه شده است استفاده کند.

تئوری واشبک

تئوری واشبک که توسط Alderson & Wall (۱۹۹۳) در مقاله "آیا واشبک وجود دارد؟" پیشنهاد شده بود شامل ۱۵ فرضیه است که در زیر آورده شده است. به غیر از آنهایی که مربوط به معلمان و تدریس می‌شود، بقیه آن‌ها به طور مستقیم یا غیرمستقیم در این مطالعه آورده شده است (فرضیه‌های ۲، ۵، ۶، ۸ و ۱۰).

۱. تست بر تدریس تأثیر می‌گذارد.
 ۲. تست بر یادگیری تأثیر می‌گذارد.
 ۳. تست بر آنچه معلم درس می‌دهد تأثیر می‌گذارد.
 ۴. تست بر چگونگی تدریس تأثیر می‌گذارد.
 ۵. تست بر آنچه دانش‌آموزان یاد می‌گیرند تأثیر می‌گذارد.
 ۶. تست بر چگونگی یادگیری تأثیر می‌گذارد.
 ۷. تست بر سرعت و ترتیب تدریس تأثیر می‌گذارد.
 ۸. تست بر سرعت و ترتیب یادگیری تأثیر می‌گذارد.
 ۹. تست بر درجه وعمق تدریس تأثیر می‌گذارد.
 ۱۰. تست بر درجه وعمق یادگیری تأثیر می‌گذارد.
 ۱۱. تست بر محتوا، روش تدریس و یادگیری تأثیر می‌گذارد.
 ۱۲. تست‌هایی که پیامدهای مهمی دارند، واشبک خواهند داشت.
 ۱۳. تست‌هایی که پیامدهای مهمی ندارند، واشبک نخواهند داشت.
 ۱۴. تست بر روی همه معلمان و دانش‌آموزان تأثیر خواهد داشت.
 ۱۵. تست تأثیر کارا بر برخی از معلمان و دانش‌آموزان می‌گذارد و بر برخی نمی‌گذارد.
- هدف Alderson & Wall (۱۹۹۳) از بیان این فرضیات برای مشخص کردن حد و مرز مطالعات واشبک و فراهم کردن موضوع برای تحقیق بوده است. سه سال بعد، Alderson and Hamp-Lyons (۱۹۹۶) فرضیه واشبک را این‌گونه تصحیح کردند: "اندازه و نوع تأثیر تست‌ها بر برخی از معلمان و دانش‌آموزان متفاوت از برخی دیگر است" (ص. ۲۹۰).

تحقیقات تجربی

تحقیقات تجربی نشان داده است که اکثر مطالعات در مورد واشبک، بر تأثیر مثبت و یا منفی امتحان بر موضوعاتی مانند محتوای کتاب، روش تدریس معلم، نگرش دانش‌آموزان و معلمان و یادگیری تأکید داشته‌اند. اختلافات زیادی در نتایج آن‌ها مشاهده شده است. بعضی‌ها واشبک مثبت را گزارش کردند و بعضی‌ها منفی را. از طرف دیگر، عده‌ایی به هر دو نوع واشبک دست یافتند در حالیکه برخی دیگر هیچ نوع واشبکی را گزارش ندادند. به بیان دیگر، نتایج تحقیقات واشبکی

که بر یادگیری دانش‌آموزان تمرکز کرده بودند بسیار پراکنده و پیچیده تر از آن بود که بتوان نتیجه قطعی گرفت.

در مطالعات گوناگون، هر دو نوع نگرش مثبت و منفی در مورد تأثیر تست گزارش شده است. به عنوان مثال، Burrows (۲۰۰۴) تأثیر یک تست کلاسی در برنامه ریزی درس زبان انگلیسی مهاجران بزرگ سال استرالیایی را مورد بررسی قرار داد. او به این نتیجه رسید که تأثیر تست بسته به نوع اعتقادات و نگرش معلمان با یکدیگر متفاوت است.

بالعکس، در برخی مطالعات تأثیری از واشبک بر میزان یادگیری دانش‌آموزان مشاهده نشد. برای مثال، Qi (۲۰۰۴) تأثیر تست انگلیسی کنکور (امتحان ورودی دانشگاه) ملی (NMET) که به منظور افزایش تأثیر مثبت بر تدریس زبان انگلیسی دبیرستان در چین بکار گرفته شده بود را بررسی کرد. نتایج تحقیق مشخص کرد که NMET تأثیر چندانی نداشته است و این تست به تنهایی نمی‌تواند تغییرات آموزشی را در مدارس چین ایجاد کند.

از طرف دیگر، اگر چه در برخی مطالعات واشبک منفی مشاهده شده بود در بیشتر تحقیقات، واشبک مثبت گزارش شده بود. به عنوان نمونه، قربانی (۲۰۰۸) تأثیر کنکور را بر نحوه تدریس معلمان پیش دانشگاهی و برنامه ریزی درسی آن‌ها را بررسی کرد. با توجه به مصاحبه‌ای که با دانش‌آموزان پیش دانشگاهی صورت گرفته بود، همگی بر تأثیر منفی کنکور تأکید داشتند و خواستار بازسازی آن بر اساس تئوری‌های کنونی تدریس و سنجش شدند. از طرف دیگر، القمدی (۲۰۰۴) تأثیر تست مهارت مرکزی ((CPT بر وضعیت آموزش و یادگیری در عربستان سعودی را بررسی کرد و دریافت که CPT تغییرات مثبتی در حیطه آموزشی عربستان ایجاد کرده است.

بعضی محققین به این نتیجه رسیده‌اند که امتحان‌های مهم و گسترده (مانند IELTS, TOEFL) تأثیر بیشتری بر محتوا و تأثیر کمتری بر روش تدریس معلمان خواهد داشت. Wall and Alderson (۱۹۹۳) مطالعه‌ای انجام دادند و تأثیر مثبت و منفی امتحان ((the O-level examination بر تدریس زبان انگلیسی را بررسی کردند. نتایج مشخص کرد که اگر چه امتحان تأثیری بر روش تدریس ندارد ولی بر محتوای تدریس تأثیر بسزایی دارد.

روش کار

شرکت کنندگان

برای انجام این تحقیق ۶۰ دانشجوی زن و مرد کارشناسی رشته آموزش زبان انگلیسی دانشگاه آزاد میناب با سطح دانش متوسط و با دامنه سنی ۲۲ تا ۳۲ سال انتخاب شدند. این تحقیق در درترم پنجم انجام گرفت. شیوه نمونه‌گیری برای انتخاب افراد، نمونه‌گیری غیر احتمالی بود و دلیل اصلی استفاده از چنین شیوه‌ای این بود که دانشگاه‌ها اجازه نمی‌دهند که کلاس‌ها به گروه‌های مختلف تفکیک شوند چراکه در این صورت حالت طبیعی به هم می‌ریزد.

ابزار

سه مجموعه ابزار ارزشیابی در این تحقیق استفاده شده است:

۱. یک آزمون استاندارد مهارت زبان (Oxford Placement Test, ۲۰۰۷) هدف استفاده از آزمون مهارت زبان، اطمینان از همسانی دانش زبانی دو گروه شرکت کننده بود. این تست توسط Lynda Edwards در دانشگاه آکسفورد در سال ۲۰۰۷ تهیه شده و شامل ۶۰ سؤال چهار گزینه‌ای لغت، گرامر و درک مطلب بود.

۲. هشت کوییز: کوییزها براساس اهداف درس توسط محقق آماده شدند و هر کدام شامل ۱۰ سؤال چهار گزینه‌ای بودند. در طول ۱۰ جلسه ۸ درس داده شد و از هر درس یک کوییز لغت گرفته شد.

۳. یک آزمون لغت: داده‌های پیش-آزمون و پس-آزمون از طریق یک تست لغت چهارگزینه‌ای که توسط محقق تهیه شده بود جمع آوری شد. این تست شامل ۴۰ سؤال بود که برای ارزیابی دانش لغوی دانش‌آموزان و بر اساس مطالبی بود که در طول کلاس تدریس شده بود (۸ درس از کتاب American Reading).

برای برطرف کردن هرگونه تفسیر اشتباه و مشکلات اجرایی، روایی آزمون لغت باید بررسی می‌شد. بنابراین، به برخی از زبان‌شناسان نشان داده شد و براساس بازخورد آن‌ها تغییراتی انجام شد تا مشکلات برطرف شود. برای بررسی روایی تست ساخته شده، با استفاده از ضریب همبستگی پیرسون (Pearson Product Correlation) نمرات تست لغت افراد با نمرات تست مهارت آن‌ها همبسته شد و در نتیجه، همبستگی قابل توجه ۰/۸۲. بین این دو تست یافت شد که روایی تست را اثبات کرد. فرمول Cronbach alpha نیز برای محاسبه اعتبار تست به کار گرفته شد و اعتبار ۰/۸۷. به دست آمد که نشانگر اعتبار تست بود.

روش

طرح پژوهشی به کار گرفته شده در این تحقیق طرح شبه تجربی است. در ابتدا برای اطمینان از همسانی افراد، یک تست مهارت زبان از همه افراد گرفته شد. سپس، یک تست لغت به عنوان پیش-آزمون از هر دو گروه آزمایش و کنترل گرفته شد تا یکسان بودن گروه‌ها از لحاظ دانش لغوی بررسی شود. پس از اطمینان از همسانی دو گروه، از گروه آزمایش ۸ کوییز گرفته شد ولی از گروه کنترل گرفته نشد. کلاس‌ها یک جلسه در هفته به مدت یک ساعت و نیم تشکیل می‌شد. برای از بین بردن تأثیر تدریس، هر دو گروه توسط یک نفر آموزش می‌دیدند. در جلسه دهم برای مشاهده تأثیر رفتار و تعیین رابطه بین متغیر وابسته (میزان فراگیری لغت) و مستقل (کوییزها)، تست لغت (پس-آزمون) بدون اطلاع قبلی به دانشجویان، از هر دو گروه گرفته شد. سرانجام، برای اطمینان از عمق یادگیری لغت که از طریق کوییزها به دست آمده بود، یک تست پس-آزمون با تأخیر از

گروه آزمایش گرفته شد.

تجزیه-تحلیل داده‌ها

فرضیه صفر: رابطه چشمگیری بین استفاده از کوییزهای متناوب لغت و میزان فراگیری دانش لغوی دانشجویان زبان انگلیسی وجود ندارد.

برای آزمودن این فرضیه، دو تی-تست نمونه‌های مستقل (independent samples t-tests) بر میانگین نمرات پیش-آزمون و پس-آزمون انجام شد. علاوه بر آن، انواع مقیاس مکرر (repeated measures ANOVA) بر رفتار گروه آزمایش در پیش-آزمون، پس-آزمون و پس-آزمون با تأخیر بکار گرفته شد تا میزان عمق یادگیری لغت دانشجویان بررسی شود و تفاوت‌های نمراتشان در این آزمون‌ها بررسی شود.

نتایج

در ابتدا برای اطمینان از همسانی افراد تی-تست نمونه‌های مستقل بر نتایج پیش-آزمون انجام شد. همانطور که در جدول مشاهده می‌کنید، تفاوت چندانی در میانگین نمرات گروه آزمایش ($M=0.5/3, SD=1.0/10$) و گروه کنترل ($M>0.5, SD=0.10/6, t_{(58)}=2.82, P=0.05$) وجود ندارد. در نتیجه، دو گروه مساویند.

Groups	N	Mean	SD	t	df	sig
Experimental group	30	10.10	3.05			
Control group	30	10.60	2.82	.65	58	.51
P < .05						

سپس، برای تعیین تأثیر کوییزهایی که فقط به گروه آزمایش داده شده بود، پس-آزمون گرفته شد. جدول زیر آمار توصیفی دو گروه در پس-آزمون را نشان می‌دهد.

Table 2. Descriptive Statistics

Groups	N	Mean	Std. Deviation	Std. Error Mean
experimental	۳۰	14.4333	2.51456	./45909
Control Post-test results	۳۰	11.0333	2.57954	./47096

به هر حال، به غیر از آمار توصیفی، تعیین اهمیت آموزش و تفاوت دقیق بین گروه‌ها، نیازمند تحلیل نتایج تی-تست نمونه‌های مستقل است که در جدول ۳ نشان داده شده است. همانطور که مشخص است، تفاوت محسوسی بین عملکرد گروه کنترل ($M=11/03, SD=2/57$) و گروه آزمایش ($M=14/43, SD=2/51; t(58)=5/17; P > .05$) دیده می‌شود. در نتیجه گروه آزمایش عملکرد بهتری نسبت به گروه کنترل داشته و فرضیه صفر رد می‌شود.

Table 3. Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig	T	Df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Post-test results Equal variances assumed	.1720	.680	5.170	58	.000	3.40000	./65770	2.08347	4.71653
Equal variances not assumed			5.170	57.962	.000	3.40000	./65770	2.08346	4.71654

اکنون که نتایج نشان داد کویزهای لغت تأثیر مثبتی بر یادگیری لغت افراد در گروه آزمایش داشته است، بررسی اینکه آیا این یادگیری با دوام بوده یا نه سودمند به نظر می‌رسد. به این منظور، انواع مقیاس مکرر بکار گرفته شد تا رفتار گروه آزمایش در پیش-آزمون، پس-آزمون و پس-آزمون با تأخیر (یک ماه بعد) بررسی شود. آمار توصیفی (جدول ۴) نشان داد که بالاترین میانگین متعلق به پس-آزمون ($M=14/43$) و کمترین آن متعلق به پیش-آزمون ($M=11/03$) است. با توجه به اینکه میانگین پس-آزمون با تأخیر ۱۳.۹۶ بوده است، مشخص شد که تفاوت زیادی بین پس-آزمون و پس-آزمون با تأخیر وجود نداشته است.

Table 4. Descriptive Statistics for Experimental Group's Performances on Pre-test (time 1), Post-test (time 2) and Delayed Post-test (time 3)

Occasion (time)	Mean	SD	N
Pre-test	10.10	3.05	30
post-test	14.43	2.51	30
Delayed post-test	13.96	2.34	30

به علاوه، نتایج انواع مقیاس مکرر مشخص کرد که تأثیر زمان قابل توجه بوده است (Wilks' Lambda = .۱۰, $F_{2/28} = 1/15$, $P < .05$, multivariate partial eta squared = .۸۹) که نشان دهنده وجود تفاوتی بین این سه نمره گروه آزمایش بوده است (جدول ۵).

Table 5. Repeated-measures ANOVA

Effect	Value	F	Hypothesis df	Error df	Sig	Partial Eta Squared
time						
Wilks' Lambda	.108	1.156E2a	2.000	28.000	.000	.892

با این وجود مشخص نشده که کدام نمرات با نمرات دیگر فرق دارد. این اطلاعات در جدول مقایسه‌ی دو به دو (جدول ۶) آورده شده است که هر دو نقطه زمانی را با هم مقایسه می‌کند و مشخص می‌کند آیا تفاوت بین آن‌ها قابل توجه است یا خیر. همانطور که در جدول مشخص شده، تفاوت بین پیش-آزمون و پس-آزمون و پیش-آزمون و پس-آزمون با تأخیر قابل توجه است ($P > .05$) درحالی‌که تفاوت بین پس-آزمون و پس-آزمون با تأخیر زیاد نیست ($P < .05$). نتایج این تحقیق نشان داد که استفاده از کویزهای لغت تأثیر زیادی بر به یادآوری لغت در دراز مدت دارد.

Table 6. Pairwise Comparisons

(I) time	(J) time	Mean Difference (I-J)	Std. Error	Sig. ^a	95% Confidence Interval for Difference ^a	
					Lower Bound	Upper Bound
1	2	-4.333*	.281	.000	-5.047	-3.619
	3	-3.867*	.306	.000	-4.644	-3.089
2	1	4.333*	.281	.000	3.619	5.047
	3	.467	.202	.084	-.046	.980
3	1	3.867*	.306	.000	3.089	4.644
	2	-.467	.202	.084	-.980	.046

Based on estimated marginal means

*The mean difference is significant at the 0/5 level.

a. Adjustment for multiple comparisons: Bonferroni.

بحث و گفتگو

بر اساس نتایج تحلیل آماری در این مطالعه، می‌توان این طور بیان کرد که تفاوت بین پیش-آزمون و پس-آزمون، تأثیر کوییزها بر یادگیری است و نتایج تی-تست مشخص کرد که تفاوت بین میانگین گروه کنترل و آزمایش قابل توجه بوده است. بنابراین فرضیه صفر رد شد. به بیان دیگر، استفاده از کوییزهای متناوب، تأثیر (واشیک) مثبت چشمگیری بر میزان یادگیری لغوی دانشجویان زبان انگلیسی داشته است و آن‌ها از ارزیابی مکرر بیشتر از ارزیابی پایانی که تنها منجر به حفظ کردن معنی لغات برای مدتی کوتاه در حافظه کوتاه مدتشان می‌شود، نفع می‌برند. برای اثبات این حرف، انواع مقیاس مکرر اجرا شد و مشخص شد که تفاوت زیادی بین میانگین گروه آزمایش در پس-آزمون و پس-آزمون با تأخیر وجود نداشته است. در حقیقت، چنین پیشرفتی در گروه کنترل دیده نشده بود. درسهای مختلف بدون دادن هیچ نوع بازخوردی از یادگیریشان به آن‌ها داده شده بود. بر طبق نتایج جدید، میتوان اینطور نتیجه گرفت که اگر از کوییزهای متناوب استفاده کنیم، یادگیری درازمدت دانش آموزان به میزان چشمگیری افزایش می‌یابد. این نتایج، از دیگر تحقیقات مانند Ari (۲۰۰۹) حمایت می‌کند که در آن تحقیق او تأثیر کوییزها بر یادگیری به عنوان یک ابزار ارزشیابی را در آمریکا مورد مطالعه قرار داده بود و به نقش مؤثر کوییزهای متناوب در یادگیری زبان آموزان پی برده بود.

حایز اهمیت است که بدانیم واشیک پدیده‌ی پیچیده‌ای است که به مکان و زمان وابسته است و از آنجایی که نتایج تحقیقات متعدد با هم همخوانی ندارند، این عقیده دوباره مورد حمایت قرار می‌گیرد. به بیان دیگر، موقعیت و مکانی که امتحان در آن برگزار می‌شود، تأثیر زیادی بر درجه و میزان واشیک دارد (Cheng et al, 2004).

کاربردهای آموزشی

اگرچه مستقیماً در این تحقیق ریشه ندارد ولی چندین کاربرد است که ممکن است این تحقیق در برداشته باشد.

کابرد برای معلمان و دانش‌آموزان

معلمان می‌توانند از نقش مثبت کوییزها در فراهم کردن بازخورد و ایجاد انگیزه بیشتر استفاده کنند. تشخیص نقاط ضعف و قوت دانش‌آموزان نیز از طریق تحلیل کوییزها به راحتی امکان پذیر است. نتایج این تحقیق می‌تواند منبع اطلاعاتی برای معلمان باشد تا بدانند چگونه می‌توانند واشبک مثبت را افزایش دهند که در نهایت منجر به بهبود یادگیری لغت دانش‌آموزان شود.

کابرد برای طرح درس نویسان و مؤلفین کتب درسی:

با علم به این که تست‌هایی که پیامدهای مهمی ندارند واشبکی هم نخواهند داشت، این تحقیق به طرح درس نویسان و مؤلفین کتب درسی کمک میکند تا یک مجموعه کوییزهای نوشته شده در کتب در نظر بگیرند و درس‌هایی را طراحی کنند تا وقت کافی برای گرفتن کوییزهای متناوب را به معلم بدهد.

References

- Alderson, J. C. & Hamp-Lyons. L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-97.
- Ari, A. (2009). The effect of quizzing on learning as a tool of assessment. *Electronic Journal of Social Sciences*. 27(8), 202-210. Retrieved Winter,2009, from www.esosder.org.
- Burrows, C. (2004). Washback in classroom-based assessment: A study of the washback effect in the Australian adult migrant English program. In Cheng, L. and Watanabe, Y. (Eds), *Washback in language testing* (pp. 113–128). Mahwah, NJ: Lawrence Erlbaum.
- Cheng, L. Y. , Watanabe, Y & Curtis, A. (Eds.). (2004). *Washback in language testing: Research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum associates.
- Edwards, L. (2007). *Oxford Solutions placement test elementary to intermediate*. Retrieved February 15, 2010, from <http://top-pdf.com/oxford-solutions-elementary.html>.
- Eraut, M. (2004). A wider perspective on assessment. *Medical Education*, 38, 803-804.
- Ghorbani, M. R. ,(2008). The Washback Effect of the University Entrance Examination on Iranian English Teachers' Curricular Planning and Instruction. *The Iranian EFL Journal Quarterly*, 2. 60-87.
- Gosa, C. M. C. (2004). *Investigating washback: A case study using student diaries*. Unpublished doctoral dissertation, Lancaster University, UK.
- Heaton, J. B. (1988). *Writing English Language Tests*. UK: Longman.
- Madsen, H. S. (1983). *Techniques in testing*. New York: Oxford University Press.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13, 241-256.
- Qi, L. (2004). Has a high-stakes test produced the intended changes? In L. Cheng , Y. Watanabe & A. Curtis (Eds). *Washback in language testing: Research contexts and methods* (pp. 3–17). Mahwah, NJ: Lawrence Erlbaum.
- Wall, D. & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10 (1), 41-69