



مقایسه عملکرد دانشجویان دکتری در دو آزمون سنتی و الکترونیکی بسندگی زبان انگلیسی

مجتبی مقصودی *

وحید رضا میرزاییان **

مریم رودباری ***

چکیده

در حال حاضر دو آزمون برای دانشجویان مقطع دکتری دانشگاه‌های دولتی و پیام نور وجود دارد که یکی توسط وزارت علوم، تحقیقات و فن آوری به صورت سنتی و دیگری توسط دانشگاه پیام نور به صورت الکترونیکی طراحی و اجرا می‌گردد. با توجه به چنین اهمیتی، پژوهش حاضر قصد مقایسه عملکرد دانشجویان در مقطع دکتری در دو آزمون فوق الذکر را دارد. جامعه آماری پژوهش حاضر، شامل کلیه دانشجویان مقطع دکتری دانشگاه پیام نور بود که در هر دو آزمون سنتی و الکترونیکی توانش زبان انگلیسی شرکت کرده بودند و با استفاده از روش نمونه گیری در دسترس، انتخاب و در این پژوهش شرکت داده شدند. ابزارهای پژوهش، ضریب همبستگی برای اطلاعات آماری و یافته‌ها حاکی از ضریب همبستگی مثبت و بالا بین نتایج آزمون سنتی و الکترونیکی ($r = 0.71$) بود. در ضمن با توجه به عدم وجود بخش شنیداری در آزمون پیام نور، امکان مقایسه این بخش نیز میسر نشد. پس نتایج بخش شنیدن در آزمون وزارت علوم با نتیجه کلی آزمون پیام نور مقایسه شد و حاکی از وجود رابطه‌ای نه چندان قوی بود ($r = 0.38$) به همین دلیل، تصمیم به مقایسه‌ی دو بخش دیگر آزمون وزارت علوم با آزمون پیام نور گرفته شد. بین عملکرد داوطلبان در بخش دستور زبان آزمون وزارت علوم و نمره نهایی پیام نور نیز رابطه چندان قوی یافت نشد ($r = 0.33$). بین عملکرد داوطلبان در بخش واژگان و درک مطلب آزمون پیام نور و نمره نهایی آزمون پیام نور، رابطه‌ای قوی‌تر از دو بخش قبلی ولی نه چندان معنادار ($r = 0.53$) وجود داشت.

واژگان کلیدی

توانش زبان انگلیسی، آزمون الکترونیکی، آزمون سنتی، دانشجویان دکتری

* استادیار، گروه زبان انگلیسی، دانشگاه فرهنگیان، تهران، ایران maghsudim@yahoo.com

** استادیار، گروه زبان انگلیسی، دانشگاه الزهراء، تهران، ایران mirzaeian@alzahra.ac.ir

*** مربی، گروه زبان انگلیسی، دانشگاه آزاد اسلامی، نکا، ایران maryamlive@yahoo.com

نویسنده مسؤول یا طرف مکاتبه: مجتبی مقصودی

مقدمه

کاربرد فن آوری در سنجش زبان، به بررسی امکان سنجش توانش زبانی زبان آموزان از طریق فن آوری اطلاعات و ارتباطات (که امروزه شامل رایانه‌ی رومیزی، رایانه‌ی روپایی، گوشی‌های هوشمند، تبلت و شبکه جهانی اینترنت نیز می‌شود) می‌پردازد (Fan & Jim 2019). این حوزه، دارای شاخه‌های متنوعی هم‌چون: آزمون انطباقی، استفاده از چند رسانه‌ای در سنجش زبان و تحلیل واکنش‌های خودکار می‌شود (Chapelle & Douglas, 2006). چپل (Chapelle, 2010) سه انگیزه مهم برای استفاده از این فن آوری در سنجش زبان را مد نظر قرار می‌دهد: افزایش بهره‌وری، شبیه بودن به آزمون سنتی و خلاقیت. منظور وی از بهره‌وری، بهبود عملکرد داوطلب در این آزمون در مقایسه‌ی با آزمون سنتی است. مراد از شبیه بودن، رعایت این نکته است که آزمون، علیرغم الکترونیکی بودن، هنوز شباهت‌های خود را با آزمون سنتی حفظ نماید و مراد از خلاقیت "امکان انتخاب و استفاده آن قسمت از زبان است که از طریق فن آوری گزینش شده و برای یک موقعیت خاص مناسب باشد." (Chapelle & Douglas, 2006, p.107).

سنجش مبتنی بر فن آوری که روزگاری تصور می‌شد می‌تواند جایگزین مناسبی برای آزمون سنتی باشد (Garrette, 1991) از دهه‌ی ۱۹۸۰ میلادی به بعد دستخوش تحولات شگرفی شده‌است (Barkaoui et al., 2013). اگرچه هنوز قسمت عمده‌ی این نوع سنجش بر تبدیل آزمون‌های زبانی سنتی به الکترونیکی متکی است لیکن این حوزه تحولات زیادی را تجربه کرده‌است (Bates et. al., 2015). در بحث‌های اخیر در حوزه‌ی سنجش، یکی از دلایل استفاده از فن آوری برای سنجش زبان، کارایی عنوان شده‌است (Crossley, et al., 2014). این یعنی یکی از مزایای مهم رایانه در سنجش زبان، نحوه‌ی ارائه تست به داوطلب، سنجش پاسخ‌های داوطلب و گزارش نمره‌است. جالب است بدانیم که مهارت صحبت کردن هنوز مورد توجه استفاده از رایانه در سنجش زبان واقع نشده‌است (Luoma, 2004).

حوزه‌ی سنجش زبان به کمک فن آوری در برگیرنده‌ی جنبه‌های مختلف سنجش زبان و استفاده از فن آوری می‌باشد. سنجش زبان به کمک فن آوری، می‌تواند به صورت خطی، انطباقی و یا نیمه انطباقی باشد (Cunnings, 2012). آزمون‌های خطی، تست‌های یکسانی را در ترتیب یکسان، به همه‌ی شرکت کنندگان در آزمون ارائه می‌کنند (Freebody, 2013). در برخی از این

تست‌ها به دواطلبان، امکان بازگشت به تست‌های قبلی داده می‌شود درحالی‌که در برخی دیگر، این امکان حذف می‌گردد. در آزمون انطباقی، تست بعدی بر اساس عملکرد داوطلب در تست قبلی ارائه می‌گردد (Frost et al., 2012). پاسخ صحیح داوطلب به یک سؤال باعث ارائه‌ی تستی دشوارتر و پاسخ غلط او باعث ارائه‌ی تستی آسان‌تر می‌گردد. با توجه به این‌که در این حالت، تست‌ها منطبق با عملکرد داوطلب به وی ارائه می‌شوند هم تعداد تست‌های به کار رفته در این آزمون، بسیار کم است و هم زمان کم‌تری برای سنجش توانش زبانی داوطلب مصرف می‌شود (Link & Cunnings, 2015).

بر خلاف تست‌های خطی که همان ادامه تفکر رایج در تست‌های سنتی است تست‌های انطباقی بر اساس نظریه واکنش به سؤال طراحی می‌شود. این نظریه بر دو فرض اساسی استوار است یکپارچگی: یعنی همه‌ی تست‌ها باید یک ویژگی را بسنجند و استقلال موضعی: یعنی پاسخ داوطلب به هر تست بایستی مستقل از تست‌های دیگر باشد (Henning et al., 1985). تست‌ها در سنجش زبان به کمک فن‌آوری را می‌توان بر حسب نوع مدل نظریه واکنش به سؤال، با استفاده از یک، دو یا سه پارامتر تدوین نمود که عبارتند از: سطح دشواری تست، قابلیت تمیز دادن بین گزینه‌ها و حدس تست (Jamieson, 2005).

با توجه به محدودیت‌های سنجش زبان به کمک فن‌آوری که معمولاً شامل هزینه‌ی زیاد، استفاده مکرر از تست، مشکلات مربوط به الگوریتم برای انتخاب تست و مشکلات رعایت چهارچوب نظریه واکنش به سؤال می‌شود تست نیمه‌انطباقی به‌عنوان راهکاری میانی پیشنهاد شده است. تست نیمه‌انطباقی بر اساس سطح گروهی از تست‌ها که آزمونک نامیده می‌شوند (Winke & Fei, 2008) و یا بر اساس سطح کل تست تعیین می‌شود که متناسب با سطح توانش داوطلب بر اساس پیش‌آزمون می‌باشد (Ockey, 2009). لازم به ذکر است که اصطلاح نیمه‌انطباقی هنوز مورد پذیرش همه پژوهشگران در این حوزه قرار نگرفته است (Winke, 2006) لیکن برخی این نوع تست را، نوع خاصی از همان تست انطباقی می‌نامند (Alderson, 2005).

تست‌های زبانی که از طریق فن‌آوری اجرا می‌شوند را می‌توان به دو گونه تست‌های مبتنی بر رایانه و تست‌های مبتنی بر اینترنت دسته‌بندی کرد (Martin, 2013). تست‌های مبتنی بر رایانه، غیر اینترنتی بوده و با استفاده از لوح فشرده و یا نرم‌افزارهایی که روی رایانه نصب می‌شود اجرا

می‌گردد (Maton, 2013). تست‌های اینترنتی، از طریق اتصال به اینترنت انجام و مورد ارزیابی قرار می‌گیرد. رور (Roever, 2006) تست‌های اینترنتی را بر حسب فن آوری به کار رفته در آن به دو گروه سطح پائین و سطح بالا دسته‌بندی می‌کند. برخی پژوهشگران (Carr, 2006) پیش‌بینی می‌کنند با توجه به گسترده شدن استفاده از اینترنت و گوشی‌های هوشمند که دائم به اینترنت متصل هستند این نوع آزمون به مرور، فراگیرترین نوع آزمون شود (Maton, 2014).

یکی از مزایای سنجش زبانی به کمک فن آوری، امکان استفاده از فرمت‌های رسانه‌ای مختلف و گنجاندن آنها در تست است (Maton, 2016). بر این اساس، تست‌های زبانی فن‌آورانه یا می‌تواند فقط دارای متن باشند (مثل آزمون پیام‌نور) و یا از انواع رسانه‌ها مثل فایل‌های صوتی و ویدئویی در آنها استفاده شود. استفاده از چندرسانه‌ای که شامل صوت، تصویر، فیلم، انیمیشن و گرافیک می‌شود اخیراً توجه پژوهشگران را به خود جلب کرده است چراکه معتقدند آزمون زبان را به واقعیت نزدیک‌تر می‌کند (Matruglio et al., 2013). لیکن برخی پژوهشگران (Carr, 2006) هشدار می‌دهند که نباید از کنار این موضوع به سادگی گذشت چراکه طراحی تست با استفاده از این عناصر، سنجش زبانی را پیچیده‌تر کرده و در نهایت ممکن است به روایی تست ضربه بزند (Plakans & Gebriel, 2012). تست‌های زبانی به کمک فن آوری را می‌توان طوری طراحی کرد تا فقط یک مهارت زبانی (مثل خواندن، نوشتن، صحبت کردن و شنیدن) و یا ترکیبی از این مهارت‌ها را بسنجد. ارزیابی مهارت‌های زبانی، توامان نشان‌دهنده ماهیت پیچیده کاربرد زبان است (Chapelle et al., 2000) و گمان می‌رود تست را به موقعیت طبیعی استفاده از زبان نزدیک‌تر کند (Ockey, 2009). برای مثال از ترکیب مهارت‌های زبانی در تافل بین‌المللی استفاده شده است تا آزمون را به مواردی که بعداً داوطلب در هنگام تحصیل در یک دانشگاه انگلیسی زبان با آنها مواجه می‌شود نزدیک‌تر کنند (Sawaki et al., 2009). بر طبق نظر پلاکانز (Plakans, 2009)، طراحی تست‌هایی که مهارت‌های زبانی را به صورت ترکیبی تست کنند بسیار دشوار است و بیش‌تر از آنها در انگلیسی برای اهداف ویژه و انگلیسی برای اهداف دانشگاهی استفاده می‌شود (Douglas, 2000). در خصوص نحوه‌ی امتیاز دهی و تصحیح آزمون زبان اینترنتی، آزمون را هم می‌توان توسط انسان وهم ماشین تصحیح کرد. در صورتی که آزمون چهارجوابی باشد که معمولاً در تست‌های شنیدن، خواندن، واژگان و دستور زبان از آن استفاده

می‌شود می‌توان به راحتی آزمون را به کمک رایانه تصحیح کرد (Weigle, 2012). البته در تست مهارت نوشتن، اگر از داوطلب خواسته شود کلمات یا عبارات خاصی را بنویسد در این حالت نیز می‌توان از رایانه برای تصحیح آزمون استفاده کرد (Alderson, 2005). ارزیابی مهارت‌های نوشتن و صحبت کردن به کمک رایانه دشوارتر است و نیاز به استفاده از تکنیک‌های پردازش زبان طبیعی دارد که در برخی از نرم افزارهای موجود به کار گرفته شده‌است (Attali & Burstein, 2006; Burstein & Chodorow, 2010; Downey et al., 2008; Chapelle & Chung, 2010; Cotos, 2011). نتایج حاصله از این سیستم‌ها هم می‌تواند وضعیت کلی داوطلب را مشخص نموده و هم به داوطلب بازخورد ارائه نماید (Burstein & Chodorow, 2010).

در خصوص نوع پاسخی که داوطلب در آزمون رایانه‌ای ارائه می‌کند دو نوع پاسخ مشخص شده‌است: پاسخ انتخابی و پاسخ تولیدی (Parshall et al., 2000). در خصوص پاسخ‌های انتخابی، داوطلب یکی از گزینه‌های ارائه شده را انتخاب می‌نماید لیکن در خصوص پاسخ‌های تولیدی، داوطلب موظف به تولید پاسخ (در حد کلمه، عبارت یا متن طولانی) می‌باشد. البته ترکیب این دو نیز ممکن است بدین حالت که مثلاً کلمات در هم ریخته‌ی یک جمله به زبان آموز داده می‌شود و از وی خواسته می‌شود آنها را به صورت یک جمله مرتب نماید.

با توجه به استفاده از نتایج دو آزمون توانش زبانی وزارت علوم و پیام نور در این پژوهش، به شرح مختصر این دو آزمون می‌پردازیم. کل اطلاعات مندرج در وب سایت رسمی آزمون توانش زبانی وزارت علوم به صورت جدول تنظیم و در جدول شماره یک درج گردیده است.

جدول ۱. قسمت‌های مختلف آزمون وزارت علوم به همراه تعداد سؤال و مدت زمان پاسخگویی

تعداد	نوع سؤال	تعداد سؤال	زمان پاسخگویی
۱	شنیدن	۳۰	۲۵
۲	دستور	۳۰	۳۰
۳	واژگان و خواندن	۴۰	۴۵
	مجموع	۱۰۰	۱۰۰

با مراجعه به سایت رسمی آزمون پیام نور، هیچ اطلاعاتی در خصوص این آزمون مشاهده نشد. با مراجعه به سایت‌های اینترنتی متفرقه نیز اطلاعاتی در خصوص قسمت‌های مختلف این آزمون به دست نیامد. در نهایت در هنگام مصاحبه با چند تن از شرکت‌کنندگان در این آزمون، اطلاعاتی در خصوص این آزمون به دست آمد که در جدول دو گنجانده شده است:

جدول ۲. اطلاعات آزمون پیام نور به روایت شرکت‌کنندگان

تعداد	نوع سؤال	تعداد سؤال	زمان پاسخگویی
۱	شنیدن	---	---
۲	دستور	۳۰	۳۰
۳	واژگان و خواندن	۷۰	۴۵
	مجموع	۱۰۰	۱۲۰

توضیح این‌که بخش درک مطلب در آزمون وزارت علوم، حاوی ۴ متن است که هر متن از ۱۰ سؤال درک مطلب و واژگان تشکیل شده است و واژگان در ارتباط با متن سؤال می‌شود. بخش واژگان و درک مطلب آزمون پیام نور حاوی ۷۰ سؤال است که از تست‌های واژگان، متون کوتاه، متون متوسط و متون طولانی تشکیل شده است. امتحان فاقد نمره منفی بوده و نمره قبولی، نصف نمره یعنی ۵۰ از ۱۰۰ می‌باشد. کلیه سؤالات در یک صفحه طراحی شده است و داوطلب با بالا و پائین کشیدن اهرم کنار صفحه، در صفحه حرکت می‌کند. پس کل وقت را می‌تواند روی تمامی سؤالات صرف کرده و پاسخ‌های خود را تغییر دهد. پس از اتمام آزمون، نمره‌ی داوطلب به صورت کلی و نه به تفکیک هر قسمت نمایش داده خواهد شد.

با توجه به نوپا بودن آموزش الکترونیکی به طور عام و انجام آزمون‌های زبانی از این طریق به طور خاص، یکی از روش‌ها برای بررسی امکان‌سنجی اجرای آزمون الکترونیکی زبان، مقایسه نتایج آزمون سنتی با آزمون الکترونیکی است. با توجه به این‌که به باور پژوهشگران، آزمون زبان دکتری دانشگاه پیام نور، تنها آزمون توانش زبانی است که به صورت الکترونیکی برگزار

می‌گردد و داوطلبانی هستند که در هر دو آزمون سنتی زبان وزارت علوم، تحقیقات و فن‌آوری (Msrt) و پیام نور (Etpnu) شرکت کرده، این پژوهش بر مقایسه این دو آزمون الکترونیکی و سنتی استوار شده است. پژوهش حاضر بر اساس فرضیات ذیل استوار گشته است:

۱. رابطه‌ای بین عملکرد شرکت کنندگان در دو آزمون وزارت علوم و پیام نور وجود دارد.
۲. با توجه به وجود بخش شنیدن در آزمون وزارت علوم و عدم وجود چنین بخشی در آزمون پیام نور، رابطه‌ای بین بخش شنیدن آزمون وزارت علوم و نمره نهایی آزمون پیام نور وجود دارد.
۳. رابطه‌ای بین بخش واژگان و درک مطلب آزمون وزارت علوم و نمره کلی آزمون پیام نور وجود دارد.

روش

همان‌طور که گفته شد هدف از انجام پژوهش حاضر، انجام مقایسه بین عملکرد دانشجویان در مقطع دکتری در دو آزمون توانش زبان انگلیسی بود لذا با توجه به محدودیت افرادی که در هر دو آزمون شرکت کرده بودند کلیه دانشجویان مقطع دکتری دانشگاه پیام نور که در هر دو آزمون الکترونیکی و سنتی توانش زبان انگلیسی شرکت کرده بودند با استفاده از روش نمونه‌گیری در دسترس انتخاب و در این پژوهش شرکت داده شدند. نمرات قسمت‌های مختلف آزمون استخراج و با استفاده از ضریب همبستگی مورد مطالعه قرار گرفت. بیش‌ترین تعداد شرکت‌کننده در این پژوهش، از حوزه علوم انسانی بود و اختلاف این حوزه با دو حوزه دیگر یعنی علوم پایه و فنی مهندسی بسیار زیاد بود. جدول سه، این اطلاعات را نمایش می‌دهد.

جدول ۳. تعداد شرکت کنندگان در پژوهش به تفکیک رشته

ردیف	رشته	تعداد
۱	انسانی	۴۱
۲	پایه	۱۲
۳	مهندسی	۷

یافته‌ها

پس از دریافت اطلاعات، اقدام به انجام آمار توصیفی (میانگین و انحراف معیار) روی اطلاعات به صورت کلی و نیز به تفکیک قسمت‌های مختلف این دو آزمون شد. جدول ۴ این اطلاعات را نشان می‌دهد.

جدول ۴، آمار توصیفی حاصل از نتایج هر دو آزمون را به تفکیک قسمت‌های مختلف نشان می‌دهد.

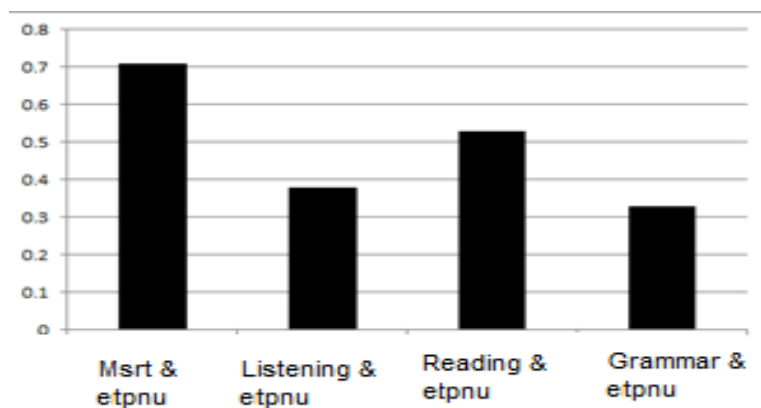
جدول ۴. آمار توصیفی نتایج هر دو آزمون

ردیف	آزمون	میانگین	انحراف معیار	حداقل	حداکثر
۱	Msrt	۵۱/۹۲	۱۰/۵	۳۲	۷۷
۲	Etpnu	۵۳/۹۳	۱۲/۳۵	۳۵	۸۷
۳	MSRT Grammar	۱۷/۶۳	۵/۵۵	۱۰	۲۹
۴	MSRT Reading	۲۲/۲۸	۸/۰۱	۱۰	۴۰
۵	MSRT Listening	۱۲	۳/۴۷	۵	۱۹

پس از کسب اطلاعات به دست آمده، اقدام به انجام محاسبات همبستگی برای آزمون فرضیه های این پژوهش شد. جدول ۵، نتایج ضریب همبستگی پیرسون برای متغیرهای پژوهش را نمایش داده و علاوه بر آن، نتایج به صورت نموداری در شکل ۱ به تصویر کشیده شده است.

جدول ۵. نتایج ضریب همبستگی پیرسون

ردیف	سؤال پژوهش	ضریب همبستگی پیرسون
۱	MSRT & ETPNU	۰/۷۱
۲	MSRT Listening & ETPNU	۰/۳۸
۳	ETPNU Reading & MSRT	۰/۵۳
۴	ETPNU Grammar & MSRT	۰/۵۳



نمودار ۱. نتایج حاصل از چهار سؤال پژوهش

بحث و نتیجه گیری

پژوهش‌های زیادی در خصوص روایی ساختاری آزمون‌های زبانی فن آورانه صورت گرفته است. منظور از روایی ساختاری، "توجیه علمی و تجربی نتایج حاصل از آزمون می‌باشد" (Chapelle, 1998). بر طبق نظر دوئی (Doey, 2008) یکی از پژوهش‌ها در این حوزه توسط

ساواکی (Sawaki, 2001) صورت گرفته است. وی یافته‌های متناقضی در این خصوص یافت و نشان داد نیاز به انجام پژوهش‌های گسترده‌ای در این حوزه وجود دارد. در عوض پژوهشگر دیگری (Sawaki, 2001) نشان داد رابطه مستقیمی بین عملکرد داوطلبان در آزمون فن آوران و سنتی شنیدن وجود دارد.

اولین نکته‌ای که در این پژوهش حاصل شد رابطه‌ی آماری نسبتاً بالا بین این دو آزمون است. همان‌طور که گفتیم آزمون پیام نور فاقد بخش شنیداری می‌باشد. لازم است بررسی‌هایی بر حسب نیاز سنجی صورت گیرد و مشخص شود آیا حذف این قسمت اقدام درستی است یا خیر؟ در ضمن می‌توان در خصوص امکان سنجی برگزاری اینترنتی آزمون وزارت علوم نیز بررسی‌هایی انجام داد. نکته‌ی بسیار حائز اهمیت دیگر، میزان درصد بسیار پایین نمره‌ی شنیدن در آزمون وزارت علوم و رابطه‌ی آماری پایین آن با نمره‌ی کلی پیام نور می‌باشد. به نظر می‌رسد بخش شنیدن، یکی از مشکلات عمده‌ی دانشجویان مقطع دکتری در بخش زبان‌آموزی است. نیاز به پژوهش‌های بیش‌تری در این زمینه وجود دارد که اولاً ریشه‌ی این مشکل را شناسایی کند و ثانیاً نیاز سنجی شود که میزان اهمیت بخش شنیداری برای دانشجویان مقطع دکتری به چه میزان است؟ اگر این مهارت، دارای اهمیتی بسیار کم و یا فاقد اهمیت برای دانشجویان مقطع دکتری است پس لزوم گنجاندن آن در آزمون وزارت علوم بی‌مورد و حذف آن توسط دانشگاه پیام نور در آزمون پیام نور آن دانشگاه کار درستی بوده است به شرطی که این تصمیم ناشی از پژوهش‌های صورت گرفته و نه ملاحظات اجرایی بوده باشد.

در خصوص بخش درک مطلب، همان‌طور که می‌دانیم درک مطلب یکی از بخش‌های بسیار مهم برای دانشجویان مقطع دکتری است. ضریب همبستگی بالای این مهارت در مقایسه با مهارت‌های دیگر حاکی از این است که دانشجویان مقطع دکتری هم‌پی به اهمیت این مهارت برده‌اند و هم در آن مسلط هستند. لیکن نکته‌ای که در اینجا مطرح می‌شود مرتبط بودن متون درک مطلب با رشته‌ی تحصیلی دانشجویان مقطع دکتری است. اقدام شایسته و بجایی که وزارت بهداشت در آزمون زبان دکتری خود انجام داده، انتخاب متون مرتبط با رشته‌های پزشکی در آزمون خود است. پس شاید یک حرکت خوب در جهت بهبود آزمون‌های توانش‌زبانی دکتری،

تفکیک این آزمون بر حسب رشته‌ی تحصیلی و ارائه متون مرتبط با رشته تحصیلی به داوطلبان این آزمون باشد.

در خصوص رابطه‌ی بین دستور زبان و نمره‌ی پیام نور، همان‌طور که گفته شد سنجش دستور زبان، آن هم به شیوه‌ی سنتی آن در حال حاضر منسوخ شده‌است و در آزمون‌های امروزی تافل و آیلتس، دستور زبان بدین شیوه ارزیابی نمی‌شوند. آزمون‌های توانش زبانی امروز، بیش‌تر به کاربردی بودن نزدیک شده‌اند. لذا لازم است در این خصوص بررسی‌های بیش‌تری صورت پذیرد و به نظر می‌رسد بر اساس نتایج این پژوهش و حذف بخش دستور زبان در آزمون‌های بین‌المللی، گنجاندن تست‌های دستور زبان آن هم بدین شیوه در آزمون‌های کنونی، جای تأمل و بررسی دارد.

و نکته‌ی آخر این که جای دو مهارت بسیار مهم در آزمون‌های توانش زبانی امروزی، هم‌چون صحبت کردن و نوشتن در دو آزمون وزارت علوم و پیام نور کاملاً خالی است. علاوه بر این که بایستی در این خصوص، پژوهش‌هایی صورت گیرد و نیازسنجی انجام شود؛ در خصوص نوشتن کاملاً مشخص است که امروزه یکی از اولویت‌های کشور در حوزه‌ی پژوهش، انتشار یافته‌های علمی به زبان انگلیسی است. اکثر دانشجویان دکتری لازم است به آن مقدار نوشتار مسلط باشند تا بتوانند پژوهش‌های خویش در مقطع دکتری را به زبان انگلیسی منتشر کنند. پس گنجاندن بخش نوشتاری در آزمون زبان دکتری به نظر ضروری می‌رسد. البته با توجه به عدم امکان تصحیح خودکار آزمون‌های گفتاری و شنیداری، گمان می‌رود عدم تمایل مسئولین آزمون‌های وزارت علوم و پیام نور به گنجاندن چنین بخش‌هایی در آزمون، مشکلات اجرایی در این حوزه باشد. پس می‌توان گفت ملاحظات اجرایی آزمون در وزارت علوم، تحقیقات و فن‌آوری و دانشگاه پیام نور آنقدر قوی و گسترده‌است که بر تمامی جنبه‌های علمی این دو آزمون سایه افکنده‌است و نیاز به بازنگری جدی دارد.

با بررسی پژوهش‌های قبلی در خصوص آزمون‌سازی، بایستی به خاطر داشته باشیم که آزمون‌ها باید دارای سه ویژگی: یعنی روایی، پایایی و قابلیت اجرایی نیز باشند. گمان می‌رود در دانشگاه پیام نور، بیش‌ترین انگیزه، سهولت اجرا و سرعت محاسبه‌ی نمره نهایی بوده‌است؛ چراکه تلاش پژوهشگران برای یافتن پژوهش‌هایی در خصوص روایی و پایایی آزمون زبان پیام

نور به جایی نرسید. شاید دانشگاه پیام نور با توجه به رسالت الکترونیکی و از راه دوری که برای خود تعریف کرده است در انجام آزمون اینترنتی زبان، پیش گام شده است لیکن باید به دو نکته توجه داشت.

اولاً "آزمون‌هایی که سرنوشت افراد را تعیین می‌کنند فوق العاده حائز اهمیت هستند و بایستی دائماً روی پایایی و روایی آنها کار کرد و بر اساس نتایج به دست آمده روی آنها تغییراتی را اعمال کرد دوماً "مقایسه‌ی آزمون تافل بین المللی قدیم و جدید نشان از حذف بخش‌هایی: مثل دستور زبان و تمایل به کاربردی‌تر کردن زبان از طریق گنجاندن صحبت کردن و نوشتن در نسخه‌های جدید بوده است در حالی که با مقایسه‌ی آزمون وزارت علوم و پیام نور در می‌یابیم یکی از مهارت‌های مهم زبانی یعنی شنیدن، در این آزمون حذف گردیده است. یافته‌ها باز حاکی از توجه بیش از حد به محدودیت‌های اجرایی تست و فدا کردن روایی و پایایی تست به بهای قابلیت اجرایی آن بوده است. مسئولین، علت حذف آزمون شنیدن را محدودیت‌های فنی ابراز می‌کنند که از دید علمی اصلاً قابل توجیه نبوده و نیست.

دیدگاه متخصصین و پژوهشگران در خصوص آزمون‌های زبانی الکترونیکی، کمی متنوع است و برخی از پژوهشگران، نگران مشکلاتی هستند که در خصوص این آزمون‌ها وجود دارد. برای مثال پژوهشگری به نام اوکی (Ockey, 2009) معتقد است آزمون‌های الکترونیکی زبانی هنوز نتوانسته‌اند آن‌طور که شایسته است به تعهدات خود عمل کنند. چال هوب دوویل (-Chalhoub, 2010) نیز متذکر می‌شود آزمون‌های الکترونیکی هنوز تحولی شگرف آن‌طور که مورد انتظار است را در حوزه آزمون زبان پدید نیاورده‌اند. در همین حال، پژوهشگران دیگری (Douglas, 2010; Douglas & Hegelheimer, 2007) نگرش مثبتی در خصوص آزمون‌های زبانی الکترونیکی دارند و معتقدند فن آوری بالاخره کار خود را خواهد کرد و در تمامی ابعاد نفوذ خواهد کرد و آزمون‌های زبانی نیز نمی‌توانند در این میان، استثنا باشند.

به هر حال هر کسی اذعان دارد که چالش‌هایی در این میان وجود دارد و واکنش داوطلبان نیز در این میان، حائز اهمیت است. این پژوهش از نقاط ضعفی رنج می‌برد که به برخی از آنها اشاره می‌گردد. فقدان وجود اطلاعات دقیق و پژوهش‌های متعدد در خصوص آزمون‌های توانش زبانی در کشور، باعث می‌شود نتوان به استاندارد بودن آنها مطمئن بود. در ضمن با توجه به عدم

دسترسی به سؤالات این آزمون‌ها، امکان بررسی روایی و پایایی این آزمون‌ها و استفاده از آنها در شرایط مشابه برای کسب اطلاعات بیش‌تر و انجام تحلیل‌های آماری در خصوص کیفیت تست‌های این آزمون‌ها میسر نیست. عدم همکاری در ارائه نمرات، پژوهشگران را مجبور ساخت از شرکت کنندگان، اطلاعات را دریافت دارند. اولاً در خصوص صحت نمره‌های ارائه شده تردید وجود دارد و ثانیاً می‌توان گفت با توجه به ماهیت داوطلبانه‌ی شرکت در این پژوهش، از نرمال بودن جمعیت آماری این پژوهش نمی‌توان اطمینان حاصل کرد. پس بایستی در تعمیم دادن نتایج این پژوهش محتاط بود.

با توجه به نوپا بودن انجام آزمون الکترونیکی و پیشگام بودن دانشگاه پیام نور در اجرای این نوع آزمون، می‌توان به بررسی جنبه‌های مختلف شرکت کنندگان در این آزمون پرداخت. موضوعاتی از قبیل نگرش شرکت کنندگان نسبت به آزمون‌های سنتی و الکترونیکی، بررسی رابطه‌ی رشته تحصیلی با آزمون‌های سنتی و الکترونیکی و جنیست شرکت کنندگان، مقایسه‌ی آزمون‌های دیگر با آزمون الکترونیکی پیام نور، طراحی و راه‌اندازی تست‌های الکترونیکی دیگر و بررسی عملکرد داوطلبان در این آزمون‌ها از جمله‌ی این موارد است.

با توجه به اهمیت واکنش شرکت کنندگان به این نوع آزمون جدید، می‌توان بر موضوعات دیگری هم‌چون میزان سواد اطلاعاتی شرکت کنندگان در آزمون وزارت علوم و تأثیر آن بر عملکرد آنان، میزان نگرش به رایانه در شرکت کنندگان آزمون الکترونیکی و عملکرد آنان در آن آزمون، میزان به کارگیری فن آوری اطلاعات و ارتباطات توسط شرکت کنندگان در آزمون الکترونیکی و عملکرد آنان در آن آزمون، میزان اضطراب رایانه‌ای شرکت کنندگان در آزمون الکترونیکی و عملکرد آنان در آن آزمون را مورد بررسی و پژوهش قرار داد.

همان‌طور که گفته شد آزمون زبان پیام نور بر خلاف آزمون وزارت علوم فاقد بخش شنیداری می‌باشد. می‌توان در این خصوص پژوهش نمود که آیا پژوهش‌های علمی در خصوص اخذ چنین تصمیماتی وجود داشته‌است یا صرفاً بر اساس ملاحظات اجرایی، چنین تصمیماتی اتخاذ شده‌است؟ در ضمن روی روایی و پایایی چنین آزمون‌هایی نیز می‌توان پژوهش‌هایی را صورت داد.

برگزاری آزمون‌های توانش زبانی یکسان برای همه رشته‌های تحصیلی بدون در نظر گرفتن تفاوت میان رشته‌های تحصیلی نیز یکی از مشکلات در این حوزه است که نیاز به بررسی و پژوهش بیشتر دارد. در ضمن با توجه به گرایش بیش‌تر دانشجویان دکتری به شرکت در آزمون وزارت علوم در مقایسه با سایر آزمون‌های مشابه، می‌توان بررسی نمود آیا دلایل این انتخاب، منشأ علمی داشته است یا خیر و یا عوامل تأثیر گذار بر این انتخاب چه بوده است؟

امکان مقایسه‌ی زیر بخش‌های مختلف دو آزمون وزارت علوم و پیام نور با یکدیگر به علت عدم گزارش قسمت‌های مختلف آزمون زبان پیام نور، میسر نشد. با توجه به نقش مهمی که بازخورد در یادگیری دارد لازم است نتیجه قسمت‌های مختلف این آزمون در دانشگاه پیام نور به تفکیک در اختیار داوطلب قرار گیرد تا بتواند اطلاعات دقیق‌تری از وضعیت توانش زبانی خود کسب نماید و از این اطلاعات بتوان برای پژوهش‌های بعدی نیز استفاده نمود.

و کلام آخر این که آزمون زبان پیام نور اصلاً از قابلیت‌های الکترونیکی هم‌چون انطباقی بودن آزمون و استفاده از سایر امکانات الکترونیکی هم‌چون چند رسانه‌ای استفاده نکرده است و حتی با حذف بخش شنیدن، نشان داده‌اند قصد استفاده از این امکانات را ندارند. طبق تقسیم بندی مطرح شده که آزمون‌های زبانی الکترونیکی را به دو دسته‌ی سطح بالا و سطح پائین دسته بندی می‌کنند می‌توان آزمون این دانشگاه را نوعی آزمون الکترونیکی سطح پائین دانست که در آن از تمامی قابلیت‌های الکترونیکی برای انجام بهتر و موثرتر آزمون استفاده نشده است ولی با توجه به این که، این گام می‌تواند اولین گام در حرکت به سمت آزمون‌های الکترونیکی واقعی باشد لذا نگارش این مقاله و مقاله‌های بعدی قطعاً به پیشرفت مؤثر در این زمینه کمک شایانی خواهد کرد. ناگفته نماند که آزمون وزارت علوم هنوز از فرمت قدیمی آزمون تافل بین‌المللی که هم‌اکنون منسوخ شده، استفاده می‌کند و شواهد حاکی از آن است که قصد الکترونیکی کردن آن حتی در سطح پائین را هم ندارند.

References

- Alderson, J. C. (2005). Diagnosing foreign language proficiency: The interface between learning and assessment. London, England: *Continuum International Publishing*. DOI: 10.1080/15434300701595637
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater v.2. *Journal of Technology, Learning, and Assessment*, 4(3), 1–31. <https://ejournals.bc.edu/index.php/jtla/article/view/1650>
- Barkaoui, K., Brooks, L., Swain, M., Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, 34, 304–324. <https://doi.org/10.1093/applin/ams046>
- Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Burstein, J., & Chodorow, M. (2010). Progress and new directions in technology for auto-mated essay evaluation. In R. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 529–38). Oxford, England: Oxford University Press. DOI: 10.1093/oxfordhb/9780195384253.013.0036
- Carr, N. T. (2006). Computer-based testing: Prospects for innovative assessment. In L. Ducate & N. Arnold (Eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching (CALICO monograph series, 5, pp. 289–312)*. San Marcos, TX: CALICO. DOI: 10.1111/j.1540-4781.2007.00639_23.x
- Chalhoub-Deville, M. (2010). Technology in standardized language assessments. In R. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (2nd ed., pp. 511–26). Oxford, England: Oxford University Press. DOI: 10.1093/oxfordhb/9780195384253.013.0035
- Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Second language acquisition and language testing interfaces* (pp. 32–70). Cambridge, England: Cambridge University Press. DOI: 10.1017/CBO9781139524711.004
- Chapelle, C. A. (2010). Technology in language testing[video]. Retrieved November 14, 2012 from <http://languagetesting.info/video/main.html>
- Chapelle, C. A., & Chung, Y.-R. (2010). The promise of NLP and speech processing technologies in language assessment. *Language Testing*, 27(3), 301–15. DOI: 10.1177/0265532210364405
- Chapelle, C. A., & Douglas, D. (2006). Assessing language through computer technology. Cambridge, England: *Cambridge University Press*. <https://www.cambridge.org/us/cambridgeenglish/catalog/teacher-training->

- development-and-research/assessing-language-through-computer-technology
- Chapelle, C., Grabe, W., & Berns, M. (2000). Communicative language proficiency: Definition and implications for TOEFL 2000. TOEFL monograph series 10. Princeton, NJ: *Educational Testing Service*.
https://www.ets.org/research/policy_research_reports/publications/report/1997/hzeu
- Coniam, D. (2006). Evaluating computer-based and paper-based versions of an *English-language listening test*. *ReCALL*, 18(2), 193–211. DOI: 10.1017/S0958344006000425
- Cotos, E. (2011). Potential of automated writing evaluation feedback. *CALICO Journal*, 28(2), 420–59. DOI: 10.11139/cj.28.2.420-459
- Crossley, S., Clevinger, A., Kim, Y. (2014). The role of lexical properties and cohesive devices in text integration and their effect on human ratings of speaking proficiency. *Language Assessment Quarterly*, 11, 250–270.
<https://doi.org/10.1080/15434303.2014.926905>
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382.
<https://doi.org/10.1177/0267658312443651>
- Dooley, P. (2008). Language testing and technology: Problems of transition to a new era. *ReCALL*, 20(1), 21–34 . DOI: 10.1017/S0958344008000311
- Douglas, D. (2000). Assessing languages for specific purposes. Cambridge, England: *Cambridge University Press*.
<https://www.amazon.com/Assessing-Languages-Specific-Cambridge-Assessment/dp/0521585430>
- Douglas, D. (2010). Understanding language testing. London, England: Hodder Education. <https://www.amazon.com/Understanding-Language-Testing-Dan-Douglas/dp/0340983434>
- Douglas, D., & Hegelheimer, V. (2007). Assessing language using computer technology. *Annual Review of Applied Linguistics*, 27, 115–32. DOI: 10.1017/S0267190508070062
- Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M., & Van Moere, A. (2008). Evaluation of the usefulness of the Versant for English test: A response. *Language Assessment Quarterly*, 5, 160–7. DOI: 10.1080/15434300801934744
- Fan, J., Jin, Y. (2019). Standards for language assessment: Demystifying university-level English placement testing in China. *Asia Pacific Journal of Education*. <https://doi.org/10.1080/02188791.2019.1706445>.
- Freebody, P. (2013). Knowledge and school talk: Intellectual accommodations to literacy? *Linguistics and Education*, 24, 4–7.
<https://doi.org/10.1016/j.linged.2012.11.004>

- Frost, K., Elder, C., Wigglesworth, G. (2012). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing*, 29(3), 345–370.
<https://doi.org/10.1177/0265532211424479>
- Garrett, N. (1991). Technology in the service of language learning: Trends and issues. *Modern Language Journal*, 75, 74–101. DOI: 10.1111/j.1540-4781.1991.tb01085.x
- Henning, G., Hudson, T., & Turner, J. (1985). Item response theory and the assumption of unidimensionality for language tests. *Language Testing*, 2(2), 141–54. DOI: 10.1177/026553228500200203
- Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228–42
www.cambridge.org/core/journals/annual-review-of-applied-linguistics/article/trends-in-computerbased-second-language-assessment/6C302FE25CCDB669A1984A6AE7B1D8E1
- Linck, J. A., Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65(S1), 185–207.
<https://doi.org/10.1111/lang.12117>
- Luoma, S. (2004). *Assessing speaking*. Cambridge language assessment series. Cambridge, England: *Cambridge University Press*.
<https://www.cambridge.org/us/cambridgeenglish/catalog/teacher-training-development-and-research/assessing-speaking/assessing-speaking-1>
- Martin, J. R. (2013). Embedded Literacy: *Knowledge as meaning*. *Linguistics and Education*, 24, 23–37
- Maton, K. (2013). Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education*, 24, 8–22.
<https://doi.org/10.1016/j.linged.2012.11.005>
- Maton, K. (2014). *Knowledge and knowers*. Towards a realist sociology of education. *Oxon and New York: Routledge*
- Maton, K. (2016). Legitimation code theory: Building knowledge about knowledge-building. In Maton, K., Hood, S., Shay, S. (Eds.), *Knowledge-building: Educational studies in Legitimation Code Theory* (pp. 1–24). *London: Routledge*
- Matruglio, E., Maton, K., Martin, J. (2013). Time travel: The role of temporality in enabling semantic waves in secondary school teaching. *Linguistics and Education*, 24, 38–49.
<https://doi.org/10.1016/j.linged.2012.11.007>
- Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal*, 93, 836–47 . DOI: 10.1111/j.1540-4781.2009.00976.x

- Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In W. J. Van der Linden & C. A. W. Glass (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 129–48). *Dordrecht, Netherlands: Kluwer*.
<https://www.springer.com/gp/book/9780792364252>
- Plakans, L. (2009a). Integrated assessment [video]. Retrieved: November 14, 2012 from <http://languageTesting.info/video/main.html>
- Plakans, L., Gebriel, A. (2012). A close investigation of source use in integrated second language writing tasks. *Assessing Writing*, 17, 18–34.
<https://doi.org/10.1016/j.asw.2011.09.002>
- Roever, C. (2006). Validation of a Web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229–56. DOI: 10.1191/0265532206lt329oa
- Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*, 5(2), 38–59. [http://dx.DOI: 10125/25127](http://dx.doi.org/10.1191/0265532206lt329oa)
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the *TOEFL Internet-based test*. *Language Testing*, 26(1), 5–30. DOI: 10.1177/0265532208097335
- Weigle, S. C., Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21, 118–133. <https://doi.org/10.1016/j.jslw.2012.03.004>
- Winke, P. (2006). Online assessment of foreign language proficiency: Meeting development, design, and delivery challenges. In S. Howell & M. Hricko (Eds.), *Online assessment and measurement: Case studies from teacher education, K-12 and corporate* (pp. 82–97). London, England: *Information Science Publishing*. DOI: 10.4018/978-1-59904-935-9.ch180