

## روایی و پایایی ارزیاب‌ها در نیمرخ ارزیابی توافقی شنیداری-ادراکی صوت

نازیلا سالاری مجد<sup>۱</sup>، سیده مریم خدای<sup>۱</sup>، میشل درینان<sup>۲</sup>، محمد کمالی<sup>۳</sup>، یونس امیری شوکی<sup>۴</sup>، نادر فلاحیان<sup>۵</sup>

<sup>۱</sup> - گروه گفتاردرمانی، دانشکده توانبخشی، دانشگاه علوم پزشکی تهران، ایران

<sup>۲</sup> - گروه فیزیولوژی پزشکی، بیمارستان Newcastle upon Tyne Freeman، بریتانیا

<sup>۳</sup> - گروه مدیریت توانبخشی، دانشکده علوم توانبخشی، دانشگاه علوم پزشکی ایران، تهران، ایران

<sup>۴</sup> - گروه گفتاردرمانی، دانشکده علوم توانبخشی، دانشگاه علوم پزشکی ایران، تهران، ایران

<sup>۵</sup> - گروه ارتوپدی فنی، دانشگاه علوم بهزیستی و توانبخشی، تهران، ایران

### چکیده

**زمینه و هدف:** ارزیابی شنیداری-ادراکی یک روش اصلی در شناسایی اختلال صوت و سنجش میزان پیشرفت درمان است. با این وجود، مطالعات داخلی معدودی به موضوع ارزیابی شنیداری-ادراکی صوت پرداخته‌اند. هدف مطالعه حاضر تدوین، بررسی روایی نسخه فارسی نیمرخ ارزیابی توافقی شنیداری-ادراکی صوت (اتشا) و پایایی ارزیاب‌ها با استفاده از این نیمرخ بود.

**روش بررسی:** روایی محتوایی کیفی نیمرخ با استفاده از آرای ۹ آسیب‌شناس گفتار و زبان متخصص در زمینه صوت و اختلالات آن و یک زبان‌شناس از طریق پرسش‌نامه به‌دست آمد. برای بررسی پایایی، نمونه‌های صوتی ۴۰ بزرگسال ۲۰ تا ۴۵ ساله دچار اختلال صوت با منشأ عصب‌زاد و اختلال عملکردی با و بدون ضایعه فیزیکی در حنجره و ۱۰ بزرگسال با همان محدوده سنی دارای صوت بهنجار ضبط شد. نمونه‌های صوتی حاوی کشش واکه و خواندن شش جمله نیمرخ اتشا بود. پایایی درونی با استفاده از آلفای کرونباخ و پایایی ارزیاب‌ها با استفاده از ضریب همبستگی پیرسون محاسبه شد.

**یافته‌ها:** نتایج حاصل از روایی محتوایی کیفی منجر به تدوین یک نیمرخ توافقی برای ارزیابی شنیداری ادراکی صوت شد. ثبات درونی نیمرخ براساس ضریب آلفای کرونباخ ۰/۹۵ بود. زمانی که نمره‌دهی نیمرخ دوبار توسط یک نفر انجام شد، ضریب همبستگی پیرسون بین ۰/۸۶ برای شدت کلی اختلال تا ۰/۴۲ برای زیربومی بود ( $p < ۰/۰۵$ ). هنگامی که ارزیابی‌کنندگان دو فرد متفاوت بودند، ضریب همبستگی پیرسون بین ۰/۸۵ برای شدت کلی اختلال تا ۰/۳۲ برای زیربومی بود ( $p < ۰/۰۵$ ).

**نتیجه‌گیری:** اتشا یک نیمرخ روا و پایا برای ارزیابی شنیداری-ادراکی صوت در بزرگسالان در زبان فارسی است.

**واژگان کلیدی:** نیمرخ ارزیابی توافقی شنیداری-ادراکی صوت، ارزیابی صوت، روایی، پایایی ارزیاب‌ها

(دریافت مقاله: ۹۲/۳/۲۹، پذیرش: ۹۲/۷/۲۸)

### مقدمه

است (۲۰۱). البته این روش دارای محدودیت‌های بسیاری است؛ برای مثال میزان تجربه درمانگر بر پایایی آن اثر می‌گذارد. اما به‌علت ویژگی مهم در دسترس بودن، در حال حاضر جایگاه خود را در ارزیابی‌های صوت به‌خوبی حفظ کرده است (۳).

نتایج ارزیابی ادراکی صوت در نیمرخ ثبت می‌شود. نیمرخ‌ها با ارائه یک چارچوب مشخص برای ارزیابی، مشخص کردن تکالیف مورد نیاز و با ارائه یک روش درجه‌بندی

ارزیابی شنیداری-ادراکی صوت، توصیف ویژگی‌های صوت صرفاً از طریق شنیدن است. ارزیابی ادراکی یک جنبه مهم در تشخیص قراردادی صوت است که امروزه روش اصلی در شناسایی اختلال صوت و سنجش میزان پیشرفت درمان محسوب می‌شود. این روش که در دسترس‌ترین و از پرکاربردترین شیوه‌های ارزیابی صوت و اختلالات آن شناخته می‌شود، از جنبه‌های مختلف مقرون به‌صرفه بوده و ویژگی برجسته آن عدم وابستگی به ابزار

معین برای ثبت نتایج توانسته‌اند پایایی ارزیابی‌های ادراکی را ارتقا دهند (۲). اولین تلاش برای طراحی نیمرخ توسط Laver (۱۹۸۰) صورت گرفت و منجر به تهیه طرح تحلیلی نیمرخ صوتی (Vocal Profile Analysis Schemas: VPAS) شد. آخرین نیمرخ، ارزیابی بالینی صوت ۲: (Clinical Voice Evaluation 2: CVE2) است که در سال ۲۰۰۷ توسط Williamson (۲۰۰۸) طراحی شد (۴۰۱). در میان نیمرخ‌ها، نیمرخ‌های ارزیابی شنیداری-ادراکی صوت CAPE-V (Consensus Auditory-Perceptual Evaluation of Voice) یکی از پرکاربردترین‌ها است که به دنبال گردهمایی متخصصان صوت از سراسر جهان، که توسط ASHA در سال ۲۰۰۳ برگزار شد، تهیه شد (۵). متخصصان بالینی از این نیمرخ برای درجه‌بندی چندین شاخص مرتبط با کیفیت صوت، شامل شدت کلی اختلال صوت (overall severity)، گرفتگی (roughness)، نفس‌آلودگی (breathiness)، تقلا (strain)، زیروبمی (pitch) و بلندی (loudness) استفاده می‌کنند. این نیمرخ شامل دو شاخص دیگر با عنوان «اظهارنظر راجع به وضعیت تشدید» و «ویژگی‌های دیگر» است. در شاخص اول، متخصص بالینی وضعیت تشدید بیمار یعنی بهنجار، پرخیشومی (hypernasality) و یا کم‌خیشومی (hyponasality) را توصیف می‌کند. در صورتی که علائم دیگری در صوت فرد ادراک شود در شاخص «ویژگی‌های دیگر» ثبت می‌شود. نتایج به دو صورت کمی و کیفی گزارش می‌شوند. همه ویژگی‌های ادراکی با استفاده از یک مقیاس ۱۰۰ میلی‌متری بینایی (Visual Analogue Scale: VAS) به صورت کمی درجه‌بندی می‌شوند. نتیجه‌گیری کیفی براساس امتیاز کمی و در قالب عبارات بهنجار، اختلال خفیف، متوسط و شدید است. تکالیف مورد استفاده در این نیمرخ شامل کشش واکه‌های /a/ و /i/، خواندن ۶ جمله و گفتار پیوسته به مدت ۲۰ ثانیه است. جمله اول شامل واکه‌های رایج زبان، جمله دوم دربردارنده شروع آسان و جمله سوم شامل هم‌خوان‌های صدا دار است. جمله چهارم برای شناسایی انسداد چاکنایی، جمله پنجم برای شناسایی خیشومی‌شدگی و جمله آخر برای ارزیابی فشار داخل دهانی

طراحی شده است. همچنین نیمرخ CAPE-V از نظر روایی در رأس نیمرخ‌های ارزیابی ادراکی صوت است (۶). پژوهش Zraick و همکاران (۲۰۱۱) روی روایی این نیمرخ نشان داد که نسبت به نیمرخ GRBAS که آن هم از نیمرخ‌های پرکاربرد در ارزیابی ادراکی است، به تغییرات حساس‌تر است و همچنین پایایی بیشتری دارد (۷). پژوهشگران کشورهای هند، پرتغال و آمریکای شمالی در مطالعات مختلف پایایی این نیمرخ و عوامل تأثیرگذار بر آن را بررسی کرده‌اند (۸-۱۰). Masaki (۲۰۰۹) نیز در مؤسسه تکنولوژی ماساچوست اقدام به تهیه نسخه کودکانه این نیمرخ کرده است (۲).

با این وجود، مطالعات داخلی معدودی از نیمرخ ارزیابی توافقی شنیداری-ادراکی صوت استفاده کرده‌اند (۱۱)، و تلاشی برای تهیه این نوع نیمرخ‌ها یا بررسی روایی و پایایی نیمرخ‌های موجود صورت نگرفته است و فقط از برگردان فارسی نیمرخ‌ها استفاده شده یا نتایج ارزیابی ادراکی بدون استفاده از نیمرخ ثبت شده است. با توجه به اهمیت ارزیابی شنیداری-ادراکی صوت و نقش نیمرخ‌ها در بهبود پایایی این شیوه ارزیابی، هدف از این پژوهش تهیه نسخه فارسی و بررسی روایی CAPE-V با عنوان نیمرخ ارزیابی توافقی شنیداری ادراکی صوت (آتسا) و بررسی پایایی (پایایی دورنی و پایایی ارزیاب‌ها) این نیمرخ بود.

### روش بررسی

به منظور تهیه نسخه فارسی CAPE-V، نیمرخ مذکور به همراه تکالیف بخش خواندن که شامل شش جمله بود طراحی شد. بنابر نتایج گردهمایی متخصصان صوت در سال ۲۰۰۳، هریک از جملات نیمرخ CAPE-V مشخصات آواشناختی و آکوستیکی ویژه‌ای دارد (۶). به این منظور ۱۲۰ جمله متناسب با اهداف نیمرخ طراحی شد. برای مثال، برای جمله شماره شش که هدف آن ارزیابی فشار داخل دهانی بود، ۲۰ جمله طراحی شد. سپس دو آسیب‌شناس گفتار و زبان و یک زبان‌شناس از بین ۱۲۰ جمله طراحی شده، ۳۰ جمله (به ازای هر جمله نیمرخ، ۵ جمله) را انتخاب کردند. ملاک انتخاب جملات ضمن در نظر گرفتن اهداف

حجره آن‌ها براساس نتایج معاینه پزشک متخصص گوش، حلق و بینی، سالم باشد؛ ۴- ویژگی‌های صوتی آن‌ها بنابر ارزیابی ادراکی انجام‌شده توسط دو آسیب‌شناس گفتار و زبان در محدودهٔ هنجار باشد و ۵- داده‌های اکوستیک آن‌ها براساس تحلیل‌های اکوستیک انجام‌شده با استفاده از نرم‌افزار Pratt 5.1.10 از نظر فرکانس پایه، شدت، آشفتگی فرکانس (jitter)، آشفتگی شدت (shimmer) و نسبت سیگنال به نویز با توجه به سن و جنس در محدودهٔ هنجار باشد (۱۲). آزمودنی‌ها تکلیف واکه را پس از محقق تکرار می‌کردند. هریک از جملات تکلیف خواندن نیز روی برگهٔ جداگانه‌ای نوشته شده بود. در صورتی که آزمودنی قادر به خواندن جملات نیمرخ نبود، پس از محقق تکرار می‌کرد و اگر قادر به تکرار جملات نبود از مطالعه کنار گذاشته می‌شد. نمونه‌های صوتی با استفاده از یک میکروفن کندانسور مدل USB AVID VOCAL STUDIO که به یک لپ‌تاپ سونی مدل VPCEE23FX متصل بود، با استفاده از نرم‌افزار PRO TOOLS در اتاق اکوستیک واقع در بخش شنوایی‌شناسی بیمارستان ضبط شد. شرایط ضبط نمونه‌های صوتی برای تمامی شرکت‌کنندگان یکسان بود. آزمودنی‌ها هر دو تکلیف واکه و خواندن را با بلندی و زیربومی عادی تولید کردند و فاصلهٔ میکروفن از دهان آزمودنی‌ها ۱۰cm بود. دو آسیب‌شناس گفتار و زبان متخصص در زمینهٔ صوت، بر مناسب بودن نمونه‌های صوتی از لحاظ توزیع انواع متنوع اختلال صوت (برحسب معیارهای ورود) و شدت‌های متفاوت اختلال صوت نظارت کردند.

با توجه به اینکه ارزیابی ادراکی صوت یک شیوهٔ ارزیابی ذهنی و مبتنی بر تجارب شنیداری-ادراکی متخصص است و در مطالعات مشابه (۸-۱۰) برای بررسی پایایی نیمرخ‌ها از روش‌های پایایی ارزیاب‌ها استفاده شده بود، در مرحلهٔ بعدی مطالعه حاضر نیز این شیوه برای بررسی پایایی نیمرخ انتخاب شد. به این منظور، نمونه‌های صوتی جمع‌آوری شده در قالب لوح‌های فشرده به همراه نسخهٔ فارسی نیمرخ اتشا و دستورالعمل استفاده از آن در اختیار محقق و یک آسیب‌شناس گفتار و زبان مجرب در زمینهٔ ارزیابی ادراکی صوت و اختلال آن قرار گرفت. برای جلوگیری از هر نوع

هر جملهٔ نیمرخ، متناسب بودن سادگی و روانی جملات با جملات نسخهٔ انگلیسی از نظر ویژگی‌های واج‌شناختی، ساختار دستوری و طول جمله بود. سپس به منظور رفع ابهام معنایی اصطلاحات به کار رفته در نیمرخ و تعیین روایی محتوایی کیفی آن، از نظرات ۱۰ صاحب‌نظر شامل آسیب‌شناسان گفتار و زبان متخصص در صوت و اختلالات آن و زبان‌شناس استفاده شد. علاوه بر تهیه و بررسی روایی محتوایی نسخهٔ فارسی نیمرخ، از بین ۳۰ جملهٔ پیشنهادی ۶ جملهٔ نهایی با توجه به آرای این صاحب‌نظران برگزیده شد.

پس از تهیهٔ نسخهٔ فارسی نیمرخ اتشا (پیوست) و تکلیف بخش خواندن آن، پایایی نیمرخ بررسی شد. به این منظور، نمونه‌های صوتی (شامل تکلیف واکه و خواندن جملات) بزرگسالان مبتلا به اختلال صوت و افراد دارای صوت بهنجار جمع‌آوری شد. این مرحله از مطالعه در درمانگاه گوش، حلق و بینی و بخش شنوایی‌شناسی بیمارستان آموزشی و تخصصی امیراعلم دانشگاه علوم پزشکی تهران انجام شد. جامعهٔ مورد مطالعه افراد مراجعه‌کننده با شکایت صوتی به بخش استروپوسکوپی درمانگاه و کارمندان و دانشجویان حاضر در بیمارستان در زمان اجرای پژوهش بودند. از جمعیت مورد نظر ۴۰ فرد مبتلا به اختلال صوت (۱۵ زن و ۲۵ مرد) و ۱۰ فرد دارای صوت هنجار (۵ زن و ۵ مرد) با محدودهٔ سنی ۲۰-۴۵ سال به روش غیراحتمالی در دسترس انتخاب شدند. معیارهای ورود برای افراد مبتلا به اختلال صوت به این ترتیب بود ۱- با شکایت صوتی به کلینیک مراجعه کرده باشند و اختلال صوتی آن‌ها توسط پزشک متخصص گوش، حلق و بینی و آسیب‌شناس گفتار و زبان بیمارستان تأیید شده باشد؛ ۲- براساس نتایج ویدئواستروپوسکوپی و به تشخیص پزشک متخصص گوش، حلق و بینی مبتلا به یکی از اختلالات صوت با منشأ عصب‌زاد یا اختلال صوت عملکردی با یا بدون ضایعه در حجره باشند و ۳- قادر به کشش واکه‌های /a/ و /i/ به مدت ۴ تا ۵ ثانیه باشند. معیارهای ورود برای افراد دارای صوت هنجار شامل ۱- در زمان اجرای مطالعه و دو هفته پیش از آن به سرماخوردگی مبتلا نشده باشند؛ ۲- در زمان اجرای مطالعه و دو هفته پیش از آن هیچ‌گونه شکایت صوتی نداشته باشند؛ ۳-

جدول ۱- شاخص‌های مرکزی و پراکندگی نمره‌دهی نوبت دوم محقق و آسیب‌شناس گفتار و زبان مجرب به افراد مبتلا به اختلال صوت براساس نیمرخ اتشا (n=۴۰)

شاخص	میانگین (انحراف معیار) نمره نوبت دوم محقق		میانگین (انحراف معیار) نمره آسیب‌شناس	
	مردان (n=۲۵)	زنان (n=۱۵)	مردان (n=۲۵)	زنان (n=۱۵)
شدت کلی اختلال	۳۳/۹۰ (۲۸/۸۱۹)	۲۳/۱۵ (۲۹/۷۱۵)	۴۷/۹۳ (۲۷/۸۴۲)	۲۵/۱۵ (۲۵/۵۲۰)
خشونت صدا	۳۵/۷۰ (۲۸/۶۴۵)	۲۳/۵۵ (۳۰/۳۶۰)	۴۴/۶۳ (۲۷/۸۵۷)	۲۳/۱۰ (۲۱/۰۶۵)
نفس‌آلودگی	۳۲/۵۳ (۲۷/۱۱۶)	۲۲/۰۵ (۲۴/۳۲۶)	۲۲ (۱۹/۸۲۵)	۱۹/۲۵ (۲۱/۶۹۹)
تقلا	۳۰/۵۷ (۳۰/۷۹۵)	۲۲/۱۰ (۳۰/۵۷۲)	۲۸/۷۰ (۲۲/۲۲۶)	۱۷/۶۰ (۲۳/۱۱۴)
زیرویمی	۵/۸۰ (۱۷/۸۲۰)	۶/۲۰ (۱۶/۵۳۹)	۱۴/۴۳ (۱۹/۱۶۷)	۱۰/۲۵ (۱۸/۹۴۰)
بلندی	۱۲/۳۷ (۲۲/۴۳۲)	۱۱/۱۵ (۲۰/۳۲۴)	۲۹/۱۷ (۲۳/۲۲۵)	۱۹/۰۵ (۲۲/۳۷۶)

سوگیری در هنگام ارزیابی ادراکی، فایل نمونه‌های صوتی اختلال‌دار و بهنجار دوباره کدگذاری و به‌صورت تصادفی در پوشه‌های جداگانه توزیع شد. برای تکمیل نیمرخ اتشا و با هدف بررسی پایایی ارزیاب‌ها، محقق و آسیب‌شناس گفتار و زبان مجرب با استفاده از گوشی AKG مدل K7 به نمونه‌های صوتی گوش کردند و برای هر نمونه صوتی یک نیمرخ را براساس دستورالعمل ارائه‌شده تکمیل کردند. آزمونگران مجاز بودند که هر نمونه صوتی را حداکثر دو بار گوش دهند. برای بررسی پایایی ارزیابی بین دو ارزیابی‌کننده یکسان، نیمرخ اتشا ۳۰ روز بعد دوباره توسط محقق تکمیل شد و برای بررسی پایایی بین دو ارزیابی‌کننده متفاوت، نتایج ارزیابی محقق با نتایج ارزیابی یک آسیب‌شناس گفتار و زبان مجرب مقایسه شد. در بررسی پایایی ارزیاب‌ها از ضریب همبستگی پیرسون استفاده شد.

### یافته‌ها

براساس نظرات ۱۰ آسیب‌شناس گفتار و زبان متخصص در زمینه صوت و اختلالات آن و زبان‌شناس، نسخه برگردان فارسی CAPE-V با عنوان نیمرخ اتشا ابهام معنایی ندارد و از روایی محتوایی کیفی برخوردار است. براساس آرای صاحب‌نظران

فوق از بین ۳۰ جمله پیشنهادی محققان، ۶ جمله نهایی انتخاب شد که بنابر توافق اکثریت، مناسب‌ترین جملات با در نظر گرفتن معیارهای ذکر شده در بخش روش بررسی بودند. یافته‌ها نشان داد ضریب آلفای کرونباخ برای بررسی ثبات درونی نیمرخ ۰/۹۵ است. همان‌طور که گفته شد در مرحله بعدی پایایی ارزیاب‌ها محاسبه شد. به این منظور، ابتدا نمراتی که شرکت‌کنندگان در نیمرخ اتشا و بنابر نظر محقق و آسیب‌شناس گفتار و زبان مجرب کسب کرده بودند، جمع‌آوری و ثبت شد. برای نمونه، شاخص‌های مرکزی و پراکندگی نمره‌دهی دوم محقق همراه با نمره‌دهی آسیب‌شناس گفتار و زبان مجرب به افراد مبتلا به اختلال صوت در جدول ۱ ارائه شده است. ابتدا ضریب همبستگی پیرسون بین نمره‌دهی اول و دوم محقق برای هریک از شاخص‌های نیمرخ اتشا محاسبه شد. به‌طور کلی، ضریب همبستگی پیرسون برای هریک از شاخص‌های نیمرخ اتشا بین مقادیر ۰/۴۲ تا ۰/۸۶ بود ( $p < ۰/۰۵$ ). همچنین ضریب تکرارپذیری (Interclass Correlation Coefficient: ICC) بین نمره‌دهی اول و دوم محقق برای هریک از شاخص‌های نیمرخ اتشا بین ۰/۸۶ (برای شدت کلی اختلال) تا ۰/۴۲ (برای زیرویمی) بود ( $p < ۰/۰۵$ ). بررسی نمره‌دهی اول و دوم محقق برحسب تکالیف مورد استفاده نشان

جدول ۲- ضریب همبستگی پیرسون برای شاخص‌های نیمرخ اتشا بین نمره‌دهی اول و دوم محقق (پایایی درون‌آزمونگر) ( $p < 0/05$ )

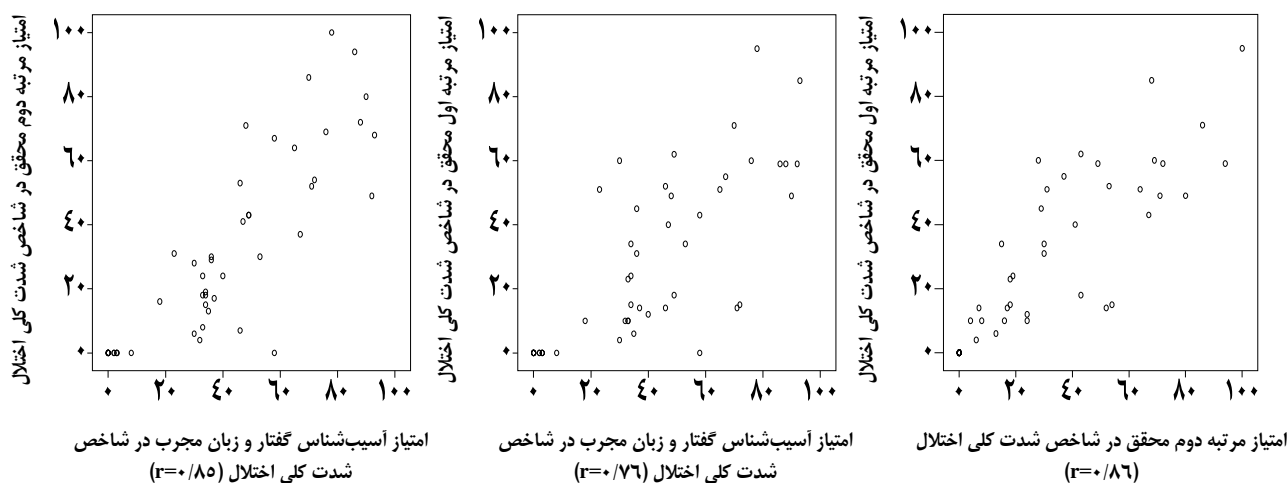
شاخص‌های نیمرخ اتشا	ضریب همبستگی در تکلیف کشش واکه	ضریب همبستگی در تکلیف خواندن	ضریب همبستگی کلی
شدت کلی اختلال	۰/۸۵	۰/۸۶	۰/۸۶
خشونت صدا	۰/۷۹	۰/۸۳	۰/۸۳
نفس‌آلودگی	۰/۷۴	۰/۷۴	۰/۷۴
تقلا	۰/۸۰	۰/۸۲	۰/۸۲
زیروبمی	۰/۴۲	۰/۴۲	۰/۴۲
بلندی	۰/۵۲	۰/۶۱	۰/۵۰

هدف از مطالعه حاضر بررسی روایی و پایایی نسخه فارسی نیمرخ CAPE-V با عنوان نیمرخ اتشا بود. آرای گروهی از آسیب‌شناسان گفتار و زبان مجرب و زبان‌شناس نشان داد نیمرخ اتشا همراه با تکالیف بخش خواندن آن روایی محتوایی کیفی دارد. بررسی یافته‌های مربوط به بخش پایایی نشان داد نیمرخ اتشا پایایی درونی مناسبی دارد. از سوی دیگر، نتایج ارزیابی ادراکی در مطالعه حاضر نشان داد کلیه شاخص‌های نیمرخ اتشا به استثنای زیروبمی و بلندی در هر دو بخش پایایی بین ارزیابی‌کنندگان یکسان و ارزیابی‌کنندگان متفاوت از قدرت تکرارپذیری مطلوبی برخوردار هستند که این موضوع بیانگر پایایی مناسب این نیمرخ است.

با در نظر گرفتن معیارهای روان‌سنجی، نیمرخ‌های روایی دارد که بتواند همان چیزی را اندازه‌گیری کند که برای آن ساخته شده است. به این ترتیب تعیین روایی محتوایی از مراحل مهم ساخت هر آزمون جدید است. Kreiman و Gerratt (۱۹۹۸) معتقدند مبحث روایی در مطالعات ارزیابی شنیداری-ادراکی صوت مورد بی‌توجهی واقع شده است و بیشتر پژوهش‌ها به بررسی پایایی این روش ارزیابی پرداخته‌اند. این مؤلفان در توجیه نظرات خود تأکید می‌کنند که روایی هر ارزیابی به پایایی آن وابسته است. بنابراین یک آزمون بدون پایایی نمی‌تواند چیزی را بسنجد که برای اندازه‌گیری آن ساخته شده است (۱۳). مطالعه Maryn و همکاران (۲۰۰۸) نشان داد استفاده از تکالیف متنوع در ارزیابی‌های

داد ضریب همبستگی پیرسون در تکلیف کشش واکه‌ها برای هریک از شاخص‌های نیمرخ اتشا بین ۰/۴۲ تا ۰/۸۵ است ( $p < 0/05$ ). در تکلیف خواندن ۶ جمله فارسی نیز ضریب همبستگی پیرسون برای هریک از شاخص‌های نیمرخ اتشا در محدوده ۰/۴۲ تا ۰/۸۶ بود ( $p < 0/05$ ) که در جدول ۲ آمده است. برای سنجش پایایی بین دو ارزیابی‌کننده متفاوت، نتایج نمره‌دهی دوم محقق با نمره‌دهی آسیب‌شناس گفتار و زبان مجرب مقایسه و ضریب همبستگی پیرسون محاسبه شد. براساس نتایج این بخش ضریب همبستگی پیرسون به‌طور کلی برای هریک از شاخص‌های نیمرخ اتشا بین ۰/۳۲ تا ۰/۸۵ بود ( $p < 0/05$ ). نمودارهای ۱ و ۲ همبستگی بین هریک از نمره‌دهی‌های محقق با آسیب‌شناس گفتار و زبان را به‌تفکیک در دو شاخص شدت کلی اختلال و زیروبمی نشان می‌دهند. جدول ۳ ضریب همبستگی پیرسون برای هریک از شاخص‌های نیمرخ در تکلیف کشش واکه‌ها را که در محدوده ۰/۳۱ تا ۰/۸۲ و در تکلیف خواندن جملات بین مقادیر ۰/۳۰ تا ۰/۸۶ بود ( $p < 0/05$ ) نشان می‌دهد. همان‌گونه که در جدول ۲ دیده می‌شود بالاترین مقدار ضریب همبستگی مربوط به شاخص شدت کلی اختلال صوت و پایین‌ترین مقدار مربوط به شاخص زیروبمی است. جدول ۳ نیز نشان می‌دهد مشابه یافته‌های فوق در مورد پایایی بین دو ارزیابی‌کننده متفاوت دیده شد.

## بحث

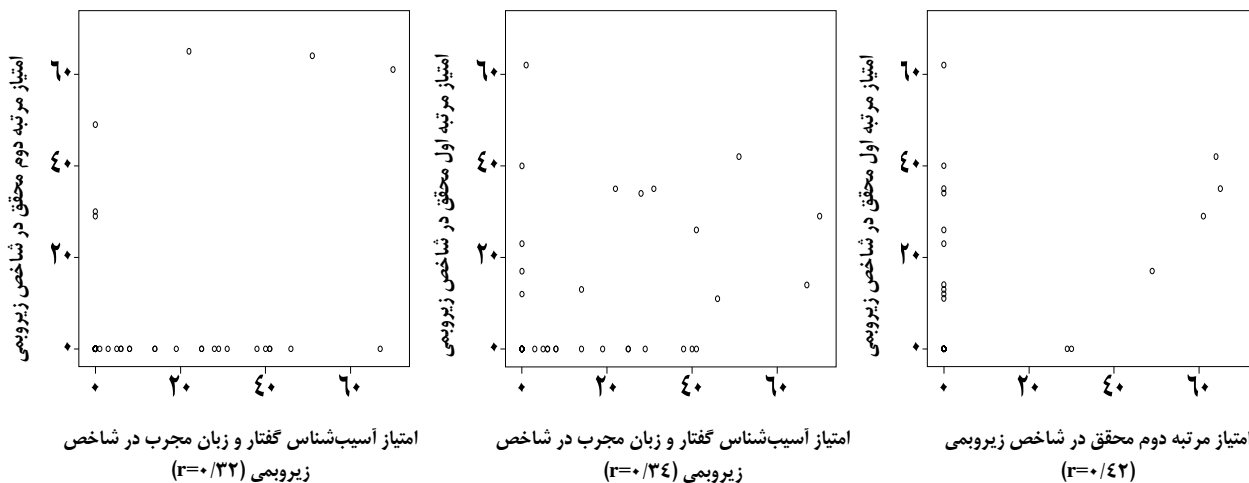


نمودار ۱- همبستگی بین امتیازهای محقق در مرتبه اول و دوم و آسیب شناس گفتار و زبان مجرب در شاخص شدت کلی اختلال از نیمرخ اتشا (n=50)

پایایی آن را بررسی کردند. آن‌ها قضاوت‌های ۲۱ متخصص مجرب صوت را در مورد ۲۲ نمونه صوت هنجار و ۳۷ نمونه مبتلا به اختلال صوت را با استفاده از دو نیمرخ CAPE-V و GRBAS جمع‌آوری کردند. نتایج پایایی بین ارزیابی‌کنندگان یکسان بین ۰/۳۵ تا ۰/۸۲ بود که بیشترین همبستگی در شاخص نفس‌آلودگی و کم‌ترین آن در شاخص تقلا ثبت شد. پایایی بین ارزیابی‌کنندگان متفاوت نیز بین ۰/۷۶ برای شاخص شدت کلی اختلال تا ۰/۲۸ برای زیربومی بود. همچنین آنها گزارش کردند که پایایی قضاوت‌های ادراکی با استفاده از CAPE-V نسبت به GRBAS به‌طور خفیفی ارتقا یافته است و این نتیجه را شواهدی برای روایی هم‌زمان CAPE-V معرفی کردند (۷). با توجه به مجموع مزایای ذکر شده، نیمرخ CAPE-V از میان نیمرخ‌های موجود انتخاب و نسخه فارسی آن تهیه شد.

پایایی یک آزمون به‌معنای تکرارپذیر بودن آن است. یک آزمون پایا می‌تواند با اطمینان در محیط‌های بالینی و پژوهشی به‌کار گرفته شود. بررسی متون نشان می‌دهد روش‌های مختلفی برای تعیین پایایی وجود دارد. در این مطالعه برای تعیین پایایی از دو روش بررسی ثبات درونی و پایایی ارزیاب‌ها استفاده شد. ضریب آلفای کرونباخ یکی از پرکاربردترین شاخص‌ها برای تعیین

شنیداری-ادراکی، سبب بهبود روایی زیست محیطی (ecological) می‌شود که اصطلاح دیگری برای اطلاق به روایی محتوایی است (۳) و این مسئله به خوبی در نیمرخ CAPE-V رعایت شده است. Kempster و همکاران (۲۰۰۹) دو نیمرخ CAPE-V و GRBAS را با یکدیگر مقایسه و این‌طور نتیجه‌گیری کردند که GRBAS به‌طور استاندارد به زبان انگلیسی منتشر نشده است و دستورالعملی در مورد نحوه استفاده از آن و تکالیف گفتاری این نیمرخ وجود ندارد. در مقابل، CAPE-V که با هدف بهبود ثبات در ارزیابی‌های بالینی بدون نیاز بیش از حد به زمان اجرا یا آموزش طراحی شده است، پروتکل ویژه‌ای در مورد طراحی تکالیف و نحوه اجرا دارد. از سوی دیگر، CAPE-V از مقیاس دیداری برای ثبت نتایج ارزیابی استفاده می‌کند، در حالی که GRBAS یک مقیاس رتبه‌ای چهار درجه‌ای دارد که به‌شدت استفاده از آن را در طراحی پروژه و تحلیل‌های آماری محدود می‌کند. آن‌ها خاطرنشان کردند که نیمرخ CAPE-V شاخص‌های صوتی بیشتری را نسبت به GRBAS ارزیابی می‌کند. این مؤلفان با در نظر گرفتن جنبه‌های فوق، CAPE-V را نیمرخ دارای روایی برای ارزیابی شنیداری-ادراکی در سال ۲۰۱۰ معرفی کردند (۶). Zraick و همکاران (۲۰۱۱) به‌منظور بررسی روایی CAPE-V،



نمودار ۲- همبستگی بین امتیازهای محقق در مرتبه اول و دوم و آسیب‌شناس گفتار و زبان مجرب در شاخص زیروبمی از نیمرخ اتشا (n=50)

باشد. درست به همین دلیل پیشنهاد می‌شود برای اطمینان از نتایج ارزیابی ادراکی زیروبمی باید مقادیر فرکانس پایه صوت نیز اندازه‌گیری شوند تا احتمال هرگونه ادراک اشتباه درباره زیروبمی کم‌تر شود. محدوده مقادیر پایایی درون‌فردی در پژوهش Ziraick و همکاران (۲۰۱۱) نیز بسیار مشابه با مطالعه حاضر بود. در هر دو پژوهش شاخص نفس‌آلودگی همبستگی بالایی داشته است، اما در مطالعه حاضر، برخلاف مطالعه Ziraick و همکاران (۲۰۱۱)، همبستگی شاخص تقلا نیز بالا بوده است.

برای تعیین پایایی بین دو ارزیابی‌کننده متفاوت ضریب همبستگی پیرسون بین نتایج ارزیابی نوبت دوم محقق و نتایج ارزیابی آسیب‌شناس مجرب گفتار و زبان محاسبه شد. نتایج نشان داد که امتیازهای به‌دست آمده از شاخص‌های شدت کلی اختلال صوت، خشونت صدا، نفس‌آلودگی و تقلا از همبستگی قوی‌ای برخوردار هستند ( $p < 0.05$ ). بدین معنا که توافق محقق و آسیب‌شناس گفتار و زبان مجرب بر سر قضاوت ادراکی درباره شاخص‌های شدت کلی اختلال صوت، خشونت صدا، نفس‌آلودگی و تقلا زیاد بود، اما نتیجه مشابه برای دو شاخص زیروبمی و بلندی دیده نشد. نتایج بررسی پایایی بین نمره‌دهی محقق و آسیب‌شناس گفتار و زبان مجرب در پژوهش حاضر، به استثنای نتایج حاصل از شاخص زیروبمی، در راستای نتایج مطالعه Prosek و Sofranko

ثبات درونی است و مقدار عددی آن بین صفر تا یک است و هرچه ضریب به عدد یک نزدیک‌تر باشد نشان می‌دهد تکرارپذیری واقعی آماری آزمون بیشتر است (۱۴). در پژوهش حاضر ضریب آلفای کرونباخ به‌دست آمده بین شش شاخص نیمرخ اتشا ۰/۹۵ بود که نشان می‌دهد نیمرخ اتشا از نظر آماری ثبات داخلی مطلوبی دارد.

در این مطالعه برای بررسی پایایی ارزیاب‌ها ضریب همبستگی پیرسون محاسبه شد. نتایج پایایی بین ارزیابی‌کنندگان یکسان نشان داد که امتیازهای ثبت‌شده توسط محقق در دو نمره‌دهی متوالی برای شاخص‌های شدت کلی اختلال صوت، خشونت صدا، نفس‌آلودگی و تقلا از همبستگی قوی برخوردار است ( $p < 0.05$ ). به عبارت دیگر، قضاوت ادراکی محقق راجع به این ۴ شاخص با وجود فاصله زمانی ۳۰ روزه به‌طور مناسب تکرار شد، اما در شاخص‌های زیروبمی و بلندی همبستگی مطلوبی دیده نشد. همسو با نتایج این بخش از مطالعه Boone و همکاران (۲۰۰۵) نیز معتقدند که ارزیابی زیروبمی به‌طور ثابت و دقیق بسیار دشوار است (۱۵). در مجموع به‌نظر می‌رسد پایین بودن مقادیر همبستگی در ارزیابی ادراکی زیروبمی چندان هم دور از انتظار نیست، زیرا ادراک زیروبمی نسبت به سایر شاخص‌های صوتی دشوارتر است؛ به‌ویژه هنگامی که در کیفیت صوت نیز اختلال

جدول ۳- ضریب همبستگی پیرسون برای شاخص‌های نیمرخ اتشا بین نمره‌دهی نوبت دوم محقق و آسیب‌شناس گفتار و زبان مجرب (پایایی بین آزمونگران) ( $p < 0.05$ )

شاخص‌های نیمرخ اتشا	ضریب همبستگی در تکلیف کشش واکه	ضریب همبستگی در تکلیف خواندن	ضریب همبستگی کلی
شدت کلی اختلال	۰/۸۲	۰/۸۶	۰/۵۸
خشونت صدا	۰/۷۰	۰/۷۷	۰/۷۹
نفس‌آلودگی	۰/۷۵	۰/۷۷	۰/۷۹
تقلا	۰/۷۲	۰/۸۰	۰/۷۹
زیرویمی	۰/۳۱	۰/۳۰	۰/۳۲
بلندی	۰/۴۲	۰/۵۸	۰/۵۲

مطالعه حاضر دو متخصص تجربه یکسانی در ارزیابی ادراکی نداشتند. نکته جالب‌توجه آن است که با وجود این تفاوت فقط توافق بر سر دو شاخص زیرویمی و بلندی نسبت به شاخص‌های دیگر کم بود. همان‌گونه که پیش از این نیز ذکر شد، نتیجه پژوهش Kempster و همکاران (۲۰۰۹) حاکی از آن بود که حساسیت نیمرخ CAPE-V به تفاوت‌های کوچک در و بین بیماران مبتلا به اختلالات صوت بیشتر است (۶) که این نتیجه می‌تواند شاهدهی دیگر بر پایایی بیشتر نیمرخ فوق باشد.

### نتیجه‌گیری

اتشا می‌تواند به‌عنوان یک نیمرخ روا و پایا برای ارزیابی شنیداری ادراکی صوت در محیط‌های بالینی و پژوهشی به‌کار رود. پایایی نتایج ارزیابی ادراکی براساس نیمرخ اتشا حتی بدون گذراندن دوره‌های آموزشی خاص برای بیشتر شاخص‌های صوتی بالا است. هرچند توافق بین ارزیابی‌کنندگان در نمره‌دهی به دو شاخص زیرویمی و بلندی نیمرخ زیاد نبود، اما بدون تردید آموزش می‌تواند میزان توافق را به مقدار قابل‌ملاحظه‌ای افزایش دهد. در مطالعات بعدی می‌توان با بررسی تأثیر عامل آموزش گام‌های مهمی برای ارتقای پایایی نیمرخ اتشا برداشت.

### سپاسگزاری

(۲۰۱۴) است که با هدف بررسی تأثیر میزان و نوع تجربه بر قضاوت‌های ادراکی صورت گرفته است. این مؤلفان نتیجه‌گیری کردند که توافق بین گفتاردرمان‌گران مجرب بسیار بالاتر از توافق گفتاردرمان‌گران بی‌تجربه بود (۱۰). باید خاطر نشان کرد که زمانی که ارزیابی توسط افراد مختلف صورت می‌گیرد تفاوت عمده بین این دو مطالعه وجود دارد، با این حال بسیاری از نتایج دو مطالعه با هم هم‌خوانی دارند که این موضوع می‌تواند شاهد دیگری بر پایایی بالای نیمرخ اتشا باشد. به‌نظر می‌رسد اگر در این پژوهش نیز مشابه مطالعه Pitts و همکاران (۲۰۰۵) نتایج ارزیابی ادراکی پس از گذراندن یک دوره آموزشی مورد بررسی قرار می‌گرفت، توافق بین ارزیابی‌کنندگان در مورد شاخص زیرویمی نیز افزایش پیدا می‌کرد. همان‌گونه که در توجیه نتایج پایایی بین دو ارزیابی‌کننده یکسان ذکر شد، دشوار بودن ادراک زیرویمی نیز می‌تواند کم‌بودن توافق بین دو آزمونگر را روی قضاوت شنیداری شاخص زیرویمی تبیین کند (۱۵).

در مطالعه‌ای دیگر، Jesus و همکاران (۲۰۰۸) مقادیر ضریب همبستگی پیرسون را برای بررسی پایایی بین دو آسیب‌شناس گفتار و زبان مجرب در کلیه شاخص‌های CAPE-V بین ۰/۸۹ تا ۰/۹۵ گزارش کردند (۹). شاید مهم‌ترین علت بالا بودن توافق بین ارزیابی‌کنندگان در پژوهش اخیر مقایسه قضاوت ادراکی دو آسیب‌شناس گفتار و زبان مجرب باشد، درحالی که در



دکتر پیمان دبیرمقدم متخصص حنجره و عضو محترم هیات علمی دانشگاه علوم پزشکی تهران، جناب آقای ابراهیمی مسئول بخش شنوایی شناسی بیمارستان امیراعلم و کلیه شرکت کنندگان در این مطالعه اعلام می‌دارند.

مقاله حاضر حاصل بخشی از پایان نامه کارشناسی ارشد در دانشگاه علوم پزشکی تهران است. نویسندگان مراتب تشکر و سپاس خود را از جناب آقای دکتر یحیی مدرسی، زبان شناس و عضو محترم هیات علمی پژوهشکده علوم انسانی، جناب آقای

## REFERENCES

- Gould J, Waugh J, Carding P, Drinnan M. A new voice rating tool for clinical practice. *J Voice*. 2012;26(4):e163-70.
- Misono S, Merati AL, Eadie TL. Developing auditory-perceptual judgment reliability in otolaryngology residents. *J Voice*. 2012;26(3):358-64.
- Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. *J Voice*. 2010;24(5):540-55.
- Webb AL, Carding PN, Deary IJ, MacKenzie K, Steen N, Wilson JA. The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol*. 2004;261(8):429-34.
- Karnell MP, Melton SD, Childes JM, Coleman TC, Dailey SA, Hoffman HT. Reliability of clinician-based (GRBAS and CAPE-V) and patient-based (V-RQOL and IPVI) documentation of voice disorders. *J Voice*. 2007;21(5):576-90.
- Kempster GB, Gerratt BR, Verdolini Abbott K, Barkmeier-Kraemer J, Hillman RE. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol*. 2009;18(2):124-32.
- Zraick RI, Kempster GB, Connor NP, Thibeault S, Klaben BK, Bursac Z, et al. Establishing validity of the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V). *Am J Speech Lang Pathol*. 2011;20(1):14-22.
- Santosh M, Rajashekhar B. Perceptual and acoustic analysis of voice in individuals with total thyroidectomy: pre-post surgery comparison [corrected]. *Indian J Otolaryngol Head Neck Surg*. 2011;63(1):32-9.
- Jesus LMT, Barney A, Couto PS, Vilarinho H, Correia A. Voice quality evaluation using CAPE-V and GRBAS in European Portuguese. *J Voice*. 2008;24(5):61-4.
- Sofranko JL, Prosek RA. The effect of levels and types of experience on judgment of synthesized voice quality. *J Voice*. 2014;28(1):24-35.
- Khoddami SM, Rabiee S, Jahani Y. Comparison of voice perceptual characteristics between speech- language pathologists', dysphonic and normal voiced adult's view. *Audiol*. 2009;18(1-2):26-35. Persian.
- Izadi F, Mohseni R, Daneshi A, Sandughdar N. Determination of fundamental frequency and voice intensity in Iranian Men and Women aged between 18 and 45 years. *J Voice*. 2012;26(3):336-40.
- Kreiman J, Gerratt BR. Validity of rating scale measures of voice quality. *J Acoust Soc Am*. 1998;104(3 Pt 1):1598-608.
- Shultz KS, Whitney DJ. Measurement theory in action: case studies and exercises. 1<sup>st</sup> ed. California: Sage Publication; 2005.
- Boone DR, McFarlane SC, Von Berg SL. The voice and voice therapy. 7<sup>th</sup> ed. Boston: Allyn & Bacon; 2004.

**Research Article**

# Validity and rater reliability of Persian version of the Consensus Auditory Perceptual Evaluation of Voice

Nazila Salary Majd<sup>1</sup>, Seyyede Maryam Khoddami<sup>1</sup>, Michael Drinnan<sup>2</sup>, Mohammad Kamali<sup>3</sup>, Yoones Amiri-Shavaki<sup>4</sup>, Nader Fallahian<sup>5</sup>

<sup>1</sup>- Department of Speech therapy, School of Rehabilitation, Tehran University of Medical Sciences, Iran

<sup>2</sup>- Department of Regional Medical Physic, Freeman Hospital, Newcastle upon Tyne, United Kingdom

<sup>3</sup>- Department of Rehabilitation Management, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran

<sup>4</sup>- Department of Speech therapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran

<sup>5</sup>- Department of Technical Orthopedic, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran

Received: 19 June 2013, accepted: 20 October 2013

## Abstract

**Background and Aim:** Auditory-perceptual assessment of voice a main approach in the diagnosis and therapy improvement of voice disorders. Despite, there are few Iranian studies about auditory-perceptual assessment of voice. The aim of present study was development and determination of validity and rater reliability of Persian version of the Consensus Auditory Perceptual Evaluation of Voice (CAPE-V).

**Methods:** The qualitative content validity was detected by collecting 10 questionnaires from 9 experienced speech and language pathologists and a linguist. For reliability purposes, the voice samples of 40 dysphonic (neurogenic, functional with and without laryngeal lesions) adults (20-45 years of age) and 10 normal healthy speakers were recorded. The samples included sustain of vowels and reading the 6 sentences of Persian version of the consensus auditory perceptual evaluation of voice called the ATSHA.

**Results:** The qualitative content validity was proved for developed Persian version of the consensus auditory perceptual evaluation of voice. Cronbach's alpha was high (0.95). Intra-rater reliability coefficients ranged from 0.86 for overall severity to 0.42 for pitch; inter-rater reliability ranged from 0.85 for overall severity to 0.32 for pitch ( $p < 0.05$ ).

**Conclusion:** The ATSHA can be used as a valid and reliable Persian scale for auditory perceptual assessment of voice in adults.

**Keywords:** Consensus Auditory Perceptual Evaluation of Voice, voice assessment, validity, rater reliability

**Please cite this paper as:** Salary Majd N, Khoddami SM, Drinnan M, Kamali M, Amiri-Shavaki Y, Fallahian N. Validity and rater reliability of Persian version of the Consensus Auditory Perceptual Evaluation of Voice. *Audiol.* 2014;23(3):65-74. Persian.