

Index for selecting an appropriate instrument to conduct research in health sciences: Introducing the COSMIN checklist

Shirinabadi Farahani A¹, *Rassouli M², Yaghmaei F³, Alavi Majd H⁴

Abstract

Introduction: Evaluating the effectiveness of health care and interventions depends on an accurate measurement, which shows the necessity of applying certain criteria for selecting an appropriate instrument. The aim of this paper was to determine the index for selecting an appropriate instrument to conduct research in health sciences.

Materials & Methods: In this review study, data were collected during 1990 till 2014 through review of Persian and English literature and purposive search for relevant index needed for selecting an appropriate instrument to conduct research in health sciences, and the most important psychometric properties of these instruments in books, articles, and databases by using the keywords: COSMIN checklist, measurement properties, psychometrics properties, disease-specific questionnaires, health status scales, health related questionnaires, psychometrics and quality criteria.

Findings: After reviewing relevant literature, 9 articles were appropriate for COSMIN as well as 24 other articles which were suitable for the most important psychometric properties of instruments. The results of this study are presented in two sections. First, the COSMIN checklist is introduced. Next, the index related to determining the psychometric properties of the instrument, i.e. validity, reliability and responsiveness are explained.

Conclusion: COSMIN checklist is an appropriate guide that facilitates the selection of a suitable instrument for researchers by providing logical index. Therefore, study results can be trusted with more confidence. The use of COSMIN Checklist in conducting research and reviewing articles in health science is recommended.

Keywords: Psychometric properties, Health sciences, Questionnaire, Checklist.

Received: 14 May 2015

Accepted: 17 June 2015

1- PhD Student in Nursing, School of Nursing and Midwifery, International Branch, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

2- Associate Professor, Department of Pediatric Nursing, School of Nursing and Midwifery, Shahid Beheshti University of Medical Sciences, Tehran, Iran. **(Correspondent author)**

E-mail: rassouli.m@gmail.com

3- Associate Professor, Department of Nursing, Zanjan Branch, Islamic Azad University, Zanjan, Iran.

4- Professor, Department of Biostatistics, School of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

شاخص های انتخاب ابزار مناسب برای انجام پژوهش در علوم سلامت: معرفی فهرست COSMIN

اعظم شیرین آبادی فراهانی^۱، *مریم رسولی^۲، فریده یغمایی^۳، حمید علوی مجد^۴

چکیده

مقدمه: ارزشیابی اثربخشی مراقبت ها و مداخلات سلامت، وابسته به اندازه گیری صحیح می باشد که این امر، ضرورت بکارگیری معیاری برای انتخاب ابزار اندازه گیری مناسب را، نشان می دهد. هدف این مقاله تعیین شاخص های انتخاب ابزارهای مناسب برای انجام پژوهش در علوم سلامت می باشد.

مواد و روش ها: در این مطالعه مروری، اطلاعات از طریق مرور متون و مقالات مرتبط و جست و جوی هدفمند منابع اطلاعاتی در زمینه ی شاخص های انتخاب ابزار مناسب پژوهش در علوم سلامت و نیز مهمترین ویژگی های روان سنجی این ابزارها در بین کتب، مقالات و بانکهای اطلاعاتی فارسی و انگلیسی، ما بین سالهای ۱۹۹۰ تا ۲۰۱۴ صورت گرفته است. جهت این امر از کلید واژه های فهرست COSMIN، ویژگیهای اندازه گیری، ویژگیهای روان سنجی، disease-specific questionnaires, health status scales, health related questionnaires, psychometrics و quality criteria استفاده گردید.

یافته ها: بعد از مرور متون مرتبط، ۹ مقاله در خصوص فهرست COSMIN و ۲۴ مقاله در خصوص مهمترین ویژگیهای روان سنجی ابزار یافت شد. یافته های این پژوهش در دو بخش مطرح گردید: نخست معرفی فهرست "استانداردهای مبتنی بر اجماع جهت انتخاب ابزارهای اندازه گیری سلامت" (Consensus-based Standards for the Selection of Health Measurement Instruments) موسوم به COSMIN و سپس توضیح در خصوص شاخص های مربوط به تعیین ویژگیهای روان سنجی ابزار یعنی روایی، پایایی و پاسخگویی آن ارایه شده است.

نتیجه گیری: فهرست COSMIN راهنمای مناسبی است که با ارائه ی شاخص های منطقی، امکان گزینش ابزار مناسب را برای پژوهشگران تسهیل نموده و در نتیجه، استناد به نتایج پژوهش ها، با اطمینان بیشتری صورت می گیرد. استفاده از فهرست فوق در انجام پژوهش و نیز داوری مقالات در علوم سلامت پیشنهاد می گردد.

کلید واژه ها: ویژگیهای روان سنجی، علوم سلامت، پرسشنامه، فهرست.

تاریخ پذیرش: ۱۳۹۴/۳/۲۷

تاریخ دریافت: ۱۳۹۴/۲/۲۴

-
- ۱- دانشجوی دکترای پرستاری، دانشکده پرستاری و مامایی، شعبه بین الملل، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.
 - ۲- دانشیار، گروه کودکان، دانشکده پرستاری و مامایی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران (نویسنده مسؤول).
پست الکترونیکی: rassouli.m@gmail.com
 - ۳- دانشیار، گروه پرستاری، دانشکده علوم پایه و پزشکی، دانشگاه آزاد اسلامی واحد زنجان، زنجان، ایران.
 - ۴- استاد، گروه آمار زیستی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران.

مقدمه

نتایج پژوهش های صورت گرفته حاکی از وجود تفاوت قابل توجه در ابعاد مختلف مراقبت بالینی بوده (۱) که همین امر موجب آشکار شدن اهمیت مراقبت مبتنی بر شواهد شده است (۲). مراقبت مبتنی بر شواهد استفاده ی آگاهانه، صحیح، آشکار و عادلانه از بهترین شواهد موجود در تصمیم گیری در مورد مراقبت از هر بیمار منحصر به فرد (۳-۵) و نیز استفاده از دقیق ترین پژوهش های موجود و ادغام آن با تجارب بالینی و ارزش های بیمار در امر مراقبت می باشد (۶). بر این اساس، پژوهشگران و دست اندرکاران مراقبت سلامت، در پی ارزیابی تاثیر روش های مراقبتی و درمانی مختلف بوده و لذا اندازه گیری صحیح، بخش مهمی از این فرایند را به خود اختصاص می دهد. به بیان دیگر، ارزشیابی اثربخشی مراقبتها یا تصمیم گیری درباره آن، وابسته به اندازه گیری صحیح می باشد (۲). سنجش و اندازه گیری که به عنوان یکی از مراحل حساس و حیاتی در پژوهش به حساب می آید، نیازمند استفاده از ابزارهای مناسب می باشد. بدین منظور، یا باید ابزاری جدید طراحی گردد و یا بعد از اطمینان یافتن از ویژگیهای روان سنجی (روایی "Validity" و پایایی "Reliability") ابزارهای خارجی، از آنها استفاده گردد (۷). قبل از بکارگیری یک ابزار، لازم است ویژگیهای روان سنجی آن ارزیابی شده و به میزان کافی مورد توجه قرار گیرد (۸-۱۰).

در طی دهه ی گذشته، مقیاسها و ابزارهای موجود در زمینه علوم سلامت به طور چشمگیری افزایش یافته و به تبع آن، مقالات زیادی به منظور جمع آوری پرسشنامه های موجود در خصوص یک مفهوم خاص و یا در جمعیت های ویژه، با هدف تعیین ویژگیهای روان سنجی ابزارهای بکار برده، به چاپ رسیده است (۲، ۱۱-۱۷). یافته مشترک این گروه از مطالعات، نشان دهنده ی آن است که ابزارهای طراحی شده، مورد قبول کاربران نبوده و در بسیاری از شرایط، قابلیت بکارگیری نداشته (۸) و یا توافقی در مورد اصطلاحات، تعاریف و ویژگیهای روان سنجی آنها دیده نمی شود (۱۸). به همین دلیل، انتخاب ابزار مناسب برای اندازه گیری متغیرهای مورد نظر، تبدیل به چالشی بزرگ شده است (۸). این امر، ضرورت به کارگیری معیارهای مناسب به منظور ارزشیابی ابزارهای موجود و شاخص های انتخاب ابزار مناسب را مورد توجه قرار می دهد. هدف این مقاله تعیین شاخص های انتخاب ابزارهای مناسب برای انجام پژوهش در علوم سلامت می باشد.

مواد و روش ها

در این مطالعه مروری، اطلاعات از طریق مرور متون و مقالات مرتبط و جست و جوی هدفمند منابع اطلاعاتی در زمینه شاخص های انتخاب ابزار مناسب پژوهش در علوم سلامت و نیز مهمترین ویژگیهای روان سنجی این ابزارها در بین کتب، مقالات و بانکهای اطلاعاتی فارسی و انگلیسی از قبیل Magiran, Google, Irandoc, Cinahl, Scholar, Proquest و Pubmed، بین سالهای ۱۹۹۰ تا ۲۰۱۴ صورت گرفته است. جهت این امر از کلید واژه های فارسی و انگلیسی فهرست COSMIN، و ویژگیهای روان سنجی (Measurement properties)، ارزیابی ویژگیهای روان سنجی، disease-specific questionnaires, health status scales, health related questionnaires, psychometrics و quality criteria استفاده گردید.

یافته ها

این قسمت، شامل دو بخش خواهد بود: از آنجا که فهرست COSMIN به عنوان مقیاسی مناسب در گزینش ابزارهای پژوهشی در علوم سلامت مطرح می باشد، لذا ابتدا به معرفی فهرست COSMIN پرداخته خواهد شد و در ادامه بخش سوم از فهرست مذکور و بررسی مهمترین ویژگیهای روان سنجی ابزار یعنی پایایی، روایی و پاسخگویی و شاخص های مربوط به تعیین هر یک از آنها ارائه می گردد.

الف- معرفی فهرست COSMIN

یکی از جامع ترین معیارها در انتخاب یک ابزار مناسب، فهرست "استانداردهای مبتنی بر اجماع جهت انتخاب ابزارهای اندازه گیری سلامت (Consensus-based Standards for the selection of health Measurement Instruments)" موسوم به COSMIN است که توسط Mokkink و همکاران و با استفاده از روش دلفی طراحی گردید (۱۰). روش دلفی، یک روش پژوهش است که بر اساس عقاید گروهی، بحث و رسیدن به اجماع، قضاوت کمی صاحب نظران، ارزیابی اولویت ها و نیز پیش بینی های دراز مدت شکل می گیرد (۱۹). در تهیه ی این فهرست، از ۷۵ نفر از متخصصین در زمینه روانشناسی، اپیدمیولوژی، آمار و پزشکی و نیز سازمانهایی نظیر "انجمن بین المللی کیفیت زندگی (International Society Quality of Life)" و "گروه پژوهش اروپایی برآیندهای سلامت (European Research Group on Health Outcomes)" که دارای کتب و

مقالاتی در خصوص ویژگیهای روان سنجی و ابزارهای اندازه گیری سلامت بودند، از نقاط مختلف جهان (شامل آمریکای شمالی، اروپا، استرالیا و آسیا)، جهت شرکت در بحث های گروهی دعوت به عمل آمد (۸، ۱۸).

فهرست طراحی شده توسط این گروه، به ارزیابی کیفیت یک پژوهش، بر اساس ویژگیهای روان سنجی ابزارهای استفاده شده در آن پژوهش، می پردازد (۸). فهرست COSMIN دارای ۱۲ حیطه مجزا است که در ۴ مرحله به بررسی ویژگیهای روان سنجی می پردازد (شکل ۱). در مرحله ی نخست، تشخیص این که کدام ویژگیها باید در یک مطالعه ارزیابی شوند، مطرح می گردد. این ویژگیها در ده حیطه بررسی می شوند. مرحله ی دوم مربوط به تشخیص این مساله است که آیا روشهای آماری استفاده شده بر اساس نظریه کلاسیک آزمون (Classical Test Theory) است یا نظریه سؤال-پاسخ (Item Response Theory). شاخصهای مربوطه، صرفاً در یک حیطه بررسی می گردد (۸، ۱۸، ۲۰). نظریه های کلاسیک آزمون و سؤال- پاسخ، در طراحی و تحلیل

پرسشنامه ها کاربرد زیادی دارند. تمرکز نظریه ی کلاسیک آزمون بیشتر بر روی عملکرد آزمون (۱۹) و نمره ای است که در آزمون کسب شده و هدف آن جستجوی ارتباط و همبستگی بین نمره واقعی و نمرات آزمون است (۲۱، ۲۲). در حالی که تمرکز نظریه سؤال- پاسخ بر عملکرد سؤال و توانایی پاسخ دهندگان (۱۹) و نحوه برخورد پاسخ دهندگان با سؤال بوده و به بررسی این موضوع می پردازد که چگونه می توان ارتباطی معنادار بین عملکرد افراد و توانایی هایی که به وسیله آزمون سنجیده می شود، برقرار نمود (۲۱، ۲۲). مرحله ی سوم، شامل بررسی روش رعایت استاندارد در تعیین ویژگیهای روان سنجی بوده که شاخصهای این مرحله در ۱۰ حیطه مجزا بررسی می شود. در این قسمت، به توصیف و ارزیابی مواردی پرداخته می شود که در مرحله ی نخست، صرفاً به آنها اشاره شده است. در نهایت در مرحله ی چهارم، الزامات عمومی برای تعمیم پذیری نتایج در یک حیطه مطرح می شود (۸، ۱۸).

جدول ۱- چهار گام تکمیل فهرست COSMIN

گام	ویژگی	حیطه
۱	تعیین ویژگیهای روان سنجی که در مطالعات مورد ارزیابی قرار گرفته اند.	روایی
		روایی محتوا
		روایی سازه
		روایی بین فرهنگی
		روایی ملاکی
		همسانی درونی
۲	تکمیل جدول مربوط در این خصوص که آیا نظریه ی سؤال- پاسخ در مطالعات استفاده شده است یا خیر.	پایایی
		خطای اندازه گیری
		پاسخگویی
		تفسیرپذیری
۳	تکمیل جدول مربوط به هر یک از ویژگیهای روان سنجی که در گام اول مشخص شد.	الزامات عمومی مطالعاتی که نظریه ی پاسخ سؤال را به کار برده اند.
۴	تکمیل جدول تعمیم پذیری برای یک از ویژگیهای روان سنجی که در گام اول مشخص شد.	تشریح ۱۰ حیطه ی مطرح شده در گام اول
		تعمیم پذیری

ب- شاخصهای مربوط به تعیین ویژگیهای روان سنجی ابزار: در میان شاخصهایی که جهت ارزیابی هر حیطه در فهرست COSMIN مطرح گشته، مواردی وجود دارد که عمومیت داشته و در تمام حیطه ها تکرار می شود نظیر: "آیا درصد داده های گمشده بیان شده است؟"، "آیا توصیفی از چگونگی مدیریت داده های گمشده وجود دارد؟"، "آیا اندازه نمونه در حیطه ی مورد نظر کافی است؟" و "آیا اشکال عمده ای در طراحی یا روش پژوهش وجود دارد؟". شاخصهای اختصاصی منتخب فهرست، در پیوست ۱ آمده است.

حیطه ی ۱۰ گانه مربوط به گام سوم فهرست COSMIN به ترتیب، عبارت از موارد ذیل است:

۱- روایی

در بررسی روایی یک ابزار، سه ویژگی روایی محتوا (Content validity)، روایی سازه (Construct validity) و روایی ملاکی (Criterion validity) لازم است مورد ارزیابی قرار گیرد (۱۸).

۱-۲ روایی محتوا

روایی بیان کننده ی این مطلب است که ابزار اندازه گیری تا چه حد ویژگی مورد نظر را می سنجد (۱۹). روایی محتوا شاخصی مهم در تشخیص مفهوم مورد اندازه گیری است و به اهدافی که یک ابزار برای رسیدن به آن طراحی شده است، اشاره دارد (۲۳). روایی صوری شاخصی ابتدایی از روایی محتوا بوده و به این سؤال پاسخ می دهد که آیا ظاهر یک ابزار برای افراد، بیماران و پاسخ دهندگان، روایی مناسبی دارد؟ (۲۴-۲۶). روش آماری خاصی برای تعیین روایی صوری وجود نداشته و به نوعی قضاوت صاحب نظران در آن دخالت دارد (۲۴). پنج شاخص در فهرست COSMIN برای ارزیابی این بخش وجود دارد که از آن جمله می توان به این موارد اشاره نمود: "آیا ارزیابی اینکه تمام گویه ها مرتبط با جامعه ی پژوهش هستند (سن، جنس، ویژگیهای بیماری، کشور و مکان)، انجام شده است؟" و "آیا ارزیابی مبنی بر این که تمام گویه ها و در ارتباط با هم، منعکس کننده ی سازه ی مورد اندازه گیری هستند، وجود دارد؟" (۸).

۲-۲ روایی سازه

روایی سازه شامل روایی ساختار (Structural validity)، آزمون فرضیه (Hypotheses testing) و روایی بین فرهنگی (Cross-cultural validity) بوده (۱۸) و به آزمون فرضیه ایجاد شده و تعیین میزان و جهت رابطه ی همبستگی میان سنجش نخستین و نتایج ایجاد شده می پردازد. اهمیت این نوع روایی به میزانی است که در صورت عدم انجام آن در یک مطالعه، احتمال

اینکه ابزار نتواند به درستی ویژگی مورد نظر را مورد سنجش قرار دهد، وجود دارد (۲۴). از روشهای متعددی برای تعیین روایی سازه استفاده می شود که از مهمترین آنها می توان به تحلیل عاملی اشاره کرد. تحلیل عاملی، روشی مناسب جهت بررسی ساختار درونی یک ابزار است (۲۷) اصولاً در طراحی ابزارهای جدید پژوهشی، تحلیل عاملی مرحله ی ارزشمندی به حساب می آید که به منظور سازمان دادن گویه ها در عوامل (زیرمقیاس ها) انجام می شود (۲۸، ۲۹). هر عامل، ویژگی خاصی را منعکس ساخته و در گروه بندی و تفسیر، راهنمای عمل پژوهشگر خواهد بود (۲۸، ۳۰) یافته های پژوهشی، در صورتی منجر به شواهد معتبر و پایا می گردند که اندازه گیری متغیرها، با استفاده از ابزارهای منطبق با فرهنگ جامعه پژوهش، صورت گیرد. واژه مبتنی بر فرهنگ، به تفاوت فرهنگی بین دو کشور یا دو گروه اشاره دارد (۱۹). اصولاً سازگار بودن یک ابزار خارجی با فرهنگ جامعه ی هدف، اصلی مهم در استفاده از این نوع ابزار بوده که با ایجاد یک زبان مشترک، مقایسه ی یافته های علمی را با سایر کشورها ممکن خواهد نمود. لذا ترجمه و مطابقت ابزار خارجی با فرهنگ جامعه ی مورد پژوهش، بسیار مهم تلقی می شود (۷، ۳۱). در میان بیست شاخص اختصاصی که به ارزیابی روایی ساختار، آزمون فرضیه و روایی بین فرهنگی می پردازند، می توان به این موارد اشاره نمود: "آیا در نظریه های کلاسیک آزمون، تحلیل عامل اکتشافی یا تاییدی انجام شده است؟"، "آیا فرضیه ای مبنی بر وجود ارتباط یا تفاوت میانگین ها قبل از جمع آوری داده ها شکل گرفته است؟"، "آیا در روایی همگرا، توصیفی کامل از ویژگیهای اندازه گیری ابزار مقایسه کننده، ارائه شده است؟"، "آیا ابزار جهت تفسیر، تطابق فرهنگی نسخه ی ترجمه شده و سهولت درک، پیش آزمون شده است؟" و "آیا عملکرد متفاوت گویه در دو گروه زبان اصلی و ترجمه شده، بررسی شده است؟" (۸).

۳-۳ روایی ملاکی

روایی ملاکی، توانایی یک ابزار در پیش بینی رفتار یک فرد در موقعیتهای ویژه تعریف شده و عملکرد هر فرد در مقایسه با یک ملاک مثل استاندارد طلایی (Gold Standard) بررسی می شود. روایی ملاکی یک ابزار، برآوردی از همبستگی بین نمرات آزمون و نمره ملاک بوده و دارای دو نوع پیش بین (predictive) و همزمان (concurrent) است (۱۹، ۲۴). هر چه میزان این همبستگی بالاتر باشد، نشاندهنده ی روایی مناسبتر ابزار جدید خواهد بود؛ هر چند گاهی یافتن ملاک مناسب بسیار دشوار خواهد بود (۳۲). هنگامی که پژوهشگر در حال بررسی ویژگی های روان سنجی ابزار جدید است،

در واقع مقیاس جدید نقش متغیر پیش بین را ایفا میکند. بنابراین، اگر نمرات ابزار جدید پیش بینی نمرات در ابزار ملاک باشد، روایی ابزار جدید، مشخص می گردد (۲۷). اگر نمرات آزمون نخست و ملاک، همزمان مورد ارزیابی قرار گیرند، روایی همزمان و اگر نمرات آزمون ملاک پس از سپری شدن یک فاصله ی زمانی از اجرای آزمون اول، مورد مطالعه قرار گیرد، روایی پیش بین صورت گرفته است (۱۹). موارد زیر شاخص های اختصاصی بررسی روایی ملاکی هستند که به عنوان نمونه آورده می شوند: "آیا ملاک به کار برده شده می تواند به عنوان یک استاندارد طلایی منطقی، محسوب گردد؟" و "آیا برای اعداد پیوسته، همبستگی یا سطح زیر منحنی عملیاتی دریافت کننده (Receiver operating curve)، محاسبه می شود؟" (۸).

۲- پایایی

در بررسی پایایی یک ابزار، عمدتاً سه ویژگی همسانی درونی (Internal consistency)، پایایی و خطای اندازه گیری (Measurement error) مورد ارزیابی قرار می گیرد (۱۸).

۱-۲ همسانی درونی

همسانی درونی به میزان همگن بودن متغیرها در درون یک ابزار اشاره داشته (۳۳) و در واقع، برآوردی از میزان همبستگی بین متغیرهایی است که سازه یا ابزار مورد نظر را تشکیل می دهند (۱۹، ۲۷). محاسبه ضریب آلفای کرونباخ (Cronbach's alpha Coefficient)، بهترین شاخص برای ارزیابی پایایی می باشد که محدوده ی صفر تا یک را در بر می گیرد. در صورتی ابزار دارای پایایی کامل است که نمره ی کسب شده، برابر یک باشد (۱۹، ۳۴). بنابر نظر Bland & Altman، ابزارهای مورد استفاده در ارزیابی های بالینی، در مقایسه با ابزارهایی که در پژوهشهای غیربالینی استفاده می شوند، باید دارای پایایی بیشتری باشند. ضریب آلفای $0.70-0.80$ در پژوهش های غیربالینی و میزانی بیش از 0.90 ($\alpha=0.90$) در پژوهشهای صورت گرفته در بالین، مناسب می باشد (۳۵). در ارزیابی همسانی درونی، علاوه بر شاخصهای عمومی، شش شاخص اختصاصی دیگر نظیر "آیا همسانی درونی برای ابزارهای تک بعدی و نیز برای هر یک از ابعاد در (ابزارهای) چند بعدی، به صورت مجزا محاسبه شده است؟" و یا "آیا در عبارات دو ارزشی، آلفای کرونباخ یا کودر ریچاردسون (Kuder Richardson ۲۰) محاسبه شده است؟" نیز در نظر گرفته شده است (۸).

۲-۲ پایایی

اندازه گیری پایایی شامل پایایی آزمون مجدد، پایایی بین مشاهده گران و پایایی درون مشاهده گران است (۱۸). پایایی ابزار،

بدان معنا است که در اندازه گیری های مختلف توسط یک ابزار، یک نتیجه یا نتایج مشابه کسب خواهد گردید (۳۲). به عبارتی، پایایی ابزار اشاره به قابلیت تکرار یا قابلیت اعتماد اندازه گیری دارد. در شرایطی که ابزار اندازه گیری پایاست، تصور بر این است که هرگونه تغییر در اندازه گیری، نشاندهنده ی تغییر واقعی در صفت مورد مطالعه است. در واقع، پایایی با این امر سر و کار دارد که ابزار اندازه گیری در شرایط یکسان تا چه اندازه نتایج یکسانی به دست می دهد (۱۹، ۳۲). پایایی آزمون مجدد (ثبات)، همسانی اندازه گیری با استفاده از یک ابزار، در دوبرابر اجرای آزمون به صورت مجزا را نشان می دهد (۱۹، ۲۷). پایایی بین مشاهده گران، به میزان توافق مشاهده گران مختلف در نمره دهی یا قضاوت در خصوص یک موقعیت یا رفتار، با همان ابزار و تحت همان معیارهای از پیش تعیین شده، اشاره دارد (۲۴). از میان ۱۰ شاخص اختصاصی که به بررسی پایایی می پردازد، مواردی چون "آیا بیماران در زمان بین انجام دو اندازه گیری جهت سنجش یک سازه، وضعیت پایداری دارند؟" و "آیا ضریب همبستگی درون طبقه ای (Intraclass Correlation Coefficient) نمرات پیوسته، محاسبه شده است؟" مطرح می باشند (۸).

۲-۳ خطای اندازه گیری

خطای معیار اندازه گیری، یکی از شاخصهای دقت اندازه گیری و پایایی آزمون و برآوردی از میزان انحراف قابل انتظار در گروهی از اندازه گیری ها در موقعیتی خاص از مقادیر واقعی، یعنی انحراف معیار توزیع نمرات خطا است (۱۹). به دلیل وجود خطا در تکرار هر اندازه گیری، همواره مقداری تفاوت و تغییر وجود دارد (۳۶، ۳۷) که در قالب "خطای معیار اندازه گیری" و "حداقل تغییر قابل تشخیص" (Minimal Detectable Changes) و به عنوان بخشی از پایایی مطلق، قابل محاسبه است (۳۸، ۳۹). این خطاها علل مختلفی دارند؛ اما به طور کلی، به دو دسته ی خطاهای سیستمی (Systematic errors) و تصادفی (Random errors) تقسیم می شوند (۲۴، ۳۶). از جمله مواردی که در قالب هفت شاخص ویژه به بررسی خطای اندازه گیری می پردازد؛ می توان به این موارد اشاره نمود: "آیا زمان فاصله گذاری بین دو آزمون گزارش شده است؟" و "آیا در نظریه های کلاسیک آزمون، خطای معیار اندازه گیری، حداقل تغییر قابل تشخیص یا محدوده ی توافق، محاسبه شده است؟" (۸).

۳- پاسخگویی (Responsiveness)

ابزارهای اندازه گیری، بایستی نسبت به تغییرات حساس بوده و قابلیت پاسخگویی داشته باشند (۴۰). پاسخگویی در دو دسته

نتیجه گیری نهایی

با توجه به آنچه گفته شد، از آنجا که امروزه استفاده از ابزارهای اندازه گیری سلامت با توجه به الزام مراقبت های مبتنی بر شواهد، رو به افزایش است. از طرفی، با توجه به طراحی مکرر این دسته از ابزارها توسط پژوهشگران مختلف در سرتاسر جهان، انتخاب و به کارگیری یک ابزار مناسب، کار دشواری به نظر می رسد. لذا لزوم استفاده از نوعی راهنما به منظور امکان انتخاب مناسب ترین ابزار، مطرح می گردد.

فهرست COSMIN راهنمای مناسبی است که با ارائه ی شاخصهای منطقی، قابل قبول و ممکن، امکان گزینش ابزار مناسب را برای پژوهشگران تسهیل نموده و با ایجاد یک استاندارد واحد، تمرکز ویژه ای را در این زمینه ایجاد کرده است. به این ترتیب هم استناد به نتایج کسب شده از پژوهش ها، با اطمینان بیشتری صورت می گیرد که منجر به ارائه ی مراقبتهای بهتر و با کیفیت در زمینه علوم سلامت می گردد و هم با تسریع در انتخاب ابزاری مناسب، از اتلاف وقت پژوهشگران جهت انتخاب ابزاری روا، پایا و پاسخگو جلوگیری می کند. پیشنهاد می گردد که پژوهشگران و داوران مقالات در حوزه ی علوم سلامت، از چک لیست فوق در انجام هر چه بهتر پژوهش و نیز دآوری مقالات استفاده نمایند.

تشکر و قدردانی

این مقاله، حاصل بخشی از رساله دکتری پرستاری اعظم شیرین آبادی فراهانی، با راهنمایی سرکار خانم دکتر مریم رسولی است. بدینوسیله از همکاری مسوولین و کارکنان دانشگاه علوم پزشکی شهید بهشتی تشکر و قدردانی می گردد.

داخلی و خارجی قرار می گیرد (۴۱). توانایی ابزار برای تشخیص تغییرات قابل اعتماد (Reliable changes) و حذف تغییرات تصادفی (Random changes) در یک بازه ی زمانی مشخص، پاسخگویی داخلی و میزان همبستگی بین تشخیص همان تغییرات در یک ابزار با ابزار ملاک و استاندارد، معرف مرتبط پاسخگویی خارجی می باشد (۲۴، ۴۱). چهارده شاخص اختصاصی برای سنجش این مورد مطرح است که به طور عمده در دو دسته قرار می گیرند: دسته نخست شامل شاخصهایی است که با وجود استاندارد طلایی، مطرح می شوند و دیگری مربوط به زمانی است که ملاک مناسبی وجود ندارد: "آیا نسبت بیماران (مثلا بر اساس میزان بهبودی یا وخامت بیماری) تغییر کرده است؟" و "آیا فرضیه ای مبنی بر تغییر نمرات، پیش از جمع آوری داده ها وجود داشته است؟" (۸).

۴- تفسیرپذیری (Interpretability)

تفسیرپذیری به معنای میزان معنی دار بودن کیفی تغییرات "حداقل تغییرات با اهمیت" (Minimal Important Changes) در نمره ی ابزار است (۱۸). حداقل تغییرات با اهمیت، کمترین تغییر در نمره ای است که نشاندهنده ی بهبودی یا وخامت وضعیت بیمار، در پاسخ به اقدام درمانی خاصی بوده و منجر به تغییر در روند مراقبت و درمان می شود. این ویژگی به درمانگر کمک می کند تا تشخیص دهد که آیا این تغییر مختصر در نمره، با استفاده از یک ابزار (مثل ابزار سنجش درد)، در بین دو گروه تحت درمان، معنی دار است و آیا تغییر در یک گروه بعد از گذشت زمان معنادار خواهد شد یا خیر (۴۲-۴۴). از بین ۹ شاخص اختصاصی جهت ارزیابی این ویژگی، سؤالاتی نظیر "آیا نمرات و تغییر نمرات (به طور مثال میانگین و انحراف معیار) برای گروه ها و زیر گروه های مرتبط (گروه های نرمال، زیر گروه بیماران یا عموم مردم)، ارائه شده است؟" و "آیا حداقل تغییر با اهمیت یا حداقل تفاوت با اهمیت، مشخص شده است؟" مطرح می باشند.

منابع

- 1- Courtney M, McCutcheon H. Using Evidence to Guide Nursing Practice. Australia: Elsevier Australia; 2010.
- 2- Gadotti I, Vieira E, Magee D. Importance and clarification of measurement properties in rehabilitation. Brazilian Journal of Physical Therapy. 2006;10(2):137-46.
- 3- Melnyk BM, Fineout-Overholt E, Stillwell SB, Williamson KM. Evidence-based practice: Step by step: The seven steps of evidence-based practice. AJN The American Journal of Nursing. 2010;110(1):51-3.
- 4- Gordon J, Watts C. Applying skills and knowledge: Principle of Nursing Practice.

Nursing Standard .Royal College of Nursing (Great Britain) : 2011; 20-26;25(33):35-7.

5- Makic MB, Martin SA, Burns S, Philbrick D, Rauen C. Putting evidence into nursing practice: Four traditional practices not supported by the evidence. *Critical Care Nurse*. 2013;33(2):28-42.

6- Houser J, Oman KS. Evidence-Base Praticie. An Implementation Guide for Healthcare Organizations. USA: Jones & Bartlett Learning, LLC.; 2011.

7- Rassouli M, Yaghmaie F, Alavi Majdm H. [Psychometric properties of "Hopefulness Scale for Adolescents" in Iranian institutionalized adolescents].

Payesh. 2010;9(2):197-204 (Persian).

8- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international delphi study. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*. 2010;19(4):539-49.

9- Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*. 2003;27(2):94-104.

10- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: Consensus-based standards for the selection of health measurement instruments. *BMC Medical Research Methodology*. 2006;6:2-9.

11-Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*. 2007;60(1):34-42.

12-Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research: An international Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*. 2012;21(4):651-7.

13- A Andresen EM. Criteria for assessing the tools of disability outcomes research. *Archives of Physical Medicine and Rehabilitation*. 2000;81(12 Suppl 2):S15-20.

14- Bombardier C, Tugwell P. Methodological considerations in functional assessment. *The Journal of Rheumatology Supplement*. 1987;14 Suppl 15:6-10.

15- Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of Neck-Specific Questionnaires: A systematic review. *BMC Medical Research Methodology*. 2011;6(11):87-101.

16- Pattanaphesaj J, Thavorncharoensap M. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. *Health and Quality of Life Outcomes*. 2015;13(1):14.

- 17- Gilchrist F, Rodd H, Deery C, Marshman Z. Assessment of the quality of measures of child oral health-related quality of life. *BMC Oral Health*. 2014;23(14): 40-57.
- 18- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*. 2010;63(7):737-45.
- 19- Waltz CF, Strickland OL, Lenz ER. *Measurement in Nursing and Health Research*. 4 th edition. New York: Springer Publishing Company, LLC; 2010.
- 20- Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Medical Research Methodology*. 2010;10: 22-30.
- 21- Ezanloo B, Habibi Asgarabadi M. Identifying differential item function (DIF) based on item-response theory: Application of the one parameter model using the BILOG MG software. *Journal of Applied Psychology*. 2010; (2(14):20-31.
- 22- Habibi M, Khodaei E, Izanloo B. [The classical and modern theories of measurement in behavioral science and medicine: A review on the methodological properties, benefits and limitations]. *Journal of Research in Behavioural Sciences*. 2012;10(4):302-15 (Persian).
- 23- Yaghmaie F. Content validity and its estimation. *Journal of Medical Education*. 2003;3(1):25-7.
- 24- McDowell II. *A Guide to Rating Scales and Questionnaires* . 3rd edition. USA: Oxford University Press, Inc; 2006.
- 25- Zamanzadeh V, Rassouli M, Abbaszadeh A, Alavi-Majd H, Nikanfar A-R, Ghahramanian A. Details of content validity and objectifying it in instrument development. *Nursing Practice Today*. 2014;1(3):163-71.
- 26- Anastasi A. *Psychological testing*. 6 ed. New York: NY: Macmillan; 1988.
- 27- Di Iorio CK. *Measurement in Health Behavior: Methods for Research and Evaluation* USA: Jossey-Bass; 2005.
- 28- Rassouli M, Zandiyeh S, Noorian M, Zayeri F, Yaghmaie F. [“Oncology Nurses Fatigue Scale”: Scale development]. *Journal of Urmia Nursing and Midwifery Faculty*. 2012;10(5):640-7 (Persian).
- 29- Munro BH. *Statistical Methods for Health Care Research*. 6th edition. Philadelphia: Lippincott Williams & Wilkins; 2009.
- 30- Gerrish K, Lacey A. *The Research Process in Nursing*. 6th edition. USA: Wiley-Blackwell 2006.
- 31- Yaghmaie F, Aminzade S, Alavi Majd H, Rassouli M, Noorian M. [Translation & evaluation of psychometric properties of the KINDL Questionnaire]. *Payesh*. 2012;11(6):841-7 (Persian).
- 32- Nunnally JC, Bernstein IH. *Psychometric theory*. USA: McGraw-Hill New York; 1994.

- 33- DeVellis RF. Scale Development. 2nd edition. London: SAGE Publications; 2003.
- 34- Deniz MS, Alsaffar AA. Assessing the validity and reliability of a questionnaire on dietary fibre-related knowledge in a Turkish student population. *Journal of Health, Population, and Nutrition*. 2013;31(4):497-503.
- 35- Sajjadi M, Rassouli M, Abbaszadeh A, Alavi Majd H, Zendehdel K. Psychometric properties of the Persian version of the Mishel's Uncertainty in Illness Scale in patients with cancer. *European Journal of Oncology Nursing : The Official Journal of European Oncology Nursing Society*. 2014;18(1):52-7.
- 36- Nordin E, Rosendahl E, Lundin-Olsson L. Timed "up & go" test: Reliability in older people dependent in activities of daily living--focus on cognitive state. *Physical Therapy*. 2006;86(5):646-55.
- 37- Zou GY. Confidence interval estimation for the Bland-Altman limits of agreement with multiple observations per individual. *Statistical Methods in Medical Research*. 2013;22(6):630-42.
- 38- Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *Journal of strength and conditioning research/ National Strength & Conditioning Association*. 2005;19(1):231-40.
- 39- Azadi F, Parnianpour M, Shaker iH, Kazemnejad A, Akbari Kamrani A, Arab A, et al. [Relative and absolute reliability of timed up and go test in community dwelling older adult and healthy young people]. *Iranian Journal of Ageing*. 2014;8(4):56-66 (Persian).
- 40- Hankins M. How discriminating are discriminative instruments? *Health and Quality of Life Outcomes*. 2008;6(36):2-5.
- 41- Jalali R. [Review of articles related to responsiveness as a characteristic of quality of life instruments]. *Journal of Knowledge & Health*. 2015;10(1):64-72 (Persian).
- 42- Khanna D, Pope JE, Khanna PP, Maloney M, Samedy N, Norrie D, et al. The minimally important difference for the Fatigue Visual Analog Scale in patients with rheumatoid arthritis followed in an academic clinical practice. *The Journal of Rheumatology*. 2008;35(12):2339-43.
- 43- Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes*. 2006;27(4):70-75.
- 44- Schuller W, Ostelo RW, Janssen R, de Vet HC. The influence of study population and definition of improvement on the smallest detectable change and the minimal important change of the Neck Disability Index. *Health and Quality of Life Outcomes*. 2014;12:53-60.

پیوست ۱- شاخصهای اختصاصی منتخب فهرست COSMIN

روایی محتوا (شامل روایی صوری): آیا ارزیابی مبنی بر اینکه تمام گویه ها به جنبه های مرتبطی از سازه مورد بررسی اشاره می کنند، انجام شده است؟ آیا ارزیابی اینکه تمام گویه ها مرتبط با جامعه پژوهش هستند (سن، جنس، ویژگیهای بیماری، کشور و مکان)، انجام شده است؟ آیا ارزیابی اینکه تمام گویه ها مرتبط با اهداف ابزار اندازه گیری هستند، انجام شده است؟ آیا ارزیابی اینکه تمام گویه ها در ارتباط با هم، منعکس کننده ی سازه ی مورد اندازه گیری هستند، انجام شده است؟	
روایی ساختار: آیا ابزار شامل شاخص های مناسب است؟ آیا در نظریه های کلاسیک آزمون، تحلیل عامل اکتشافی یا تاییدی انجام شده است؟ آیا در نظریه ی سؤال - پاسخ، آزمون مرتبط برای تشخیص بعد یا ابعاد گویه ها انجام شده است؟	۵.۴ ۵.۵
آزمون فرضیه: آیا فرضیه ای مبنی بر وجود ارتباط یا تفاوت میانگین ها قبل از جمع آوری داده ها شکل گرفته است؟ آیا فرضیه شامل جهت مورد انتظار در همبستگی یا تفاوت میانگین ها می باشد؟ آیا فرضیه شامل میزان مطلق یا نسبی مورد انتظار همبستگی یا تفاوت میانگین ها می باشد؟ آیا در روایی همگرا، توصیفی کامل از ابزار معیار، ارائه شده است؟	
روایی مبتنی بر فرهنگ: آیا تخصص هر یک از افرادی که در فرایند ترجمه شرکت داشتند، به طور کامل توصیف شده است؟ آیا مترجمین به صورت مستقل از هم کار کرده اند؟ کدام گویه ها ترجمه و ترجمه برگردان شده است؟ آیا به اندازه کافی در خصوص چگونگی رفع تفاوت بین نسخه اصلی و ترجمه شده، توضیح داده شده است؟ آیا ترجمه توسط گروهی (به عنوان مثال طراحان اولیه ابزار) مورد بازبینی قرار گرفته است؟ آیا ابزار جهت تفسیر، تطابق فرهنگی نسخه ی ترجمه شده و سهولت درک، آزمون مقدماتی انجام شده است؟ آیا نمونه ها به غیر از زبان و بستر فرهنگی در تمام ویژگیها، مشابه بودند؟ آیا عملکرد متفاوت گویه در دو گروه زبان اصلی و ترجمه شده، بررسی شده است؟	
روایی ملاکی: آیا ملاک به کار برده شده می تواند به عنوان یک استاندارد طلایی منطقی، محسوب می گردد؟ آیا برای نمرات دو ارزشی، حساسیت و ویژگی تعیین شده است؟	
همسانی درونی: آیا ابزار شامل نشانگرهای مناسب است؟ آیا تک بعدی بودن ابزار چک شده است؟ به طور مثال آیا تحلیل عاملی انجام شده است؟ آیا همسانی درونی برای ابزار تک بعدی و نیز برای هر یک از ابعاد در (ابزار) چند بعدی، به صورت مجزا محاسبه شده است؟ آیا در نظریه های کلاسیک آزمون، آلفای کرونباخ محاسبه شده است؟ آیا در عبارات سؤالات دوازدهی، آلفای کرونباخ یا کودر ریچاردسون ۲۰ محاسبه شده است؟ آیا در نظریه سؤال - پاسخ، نیکویی برازش انجام شده است؟	
پایایی: آیا حداقل دو اندازه گیری وجود دارد؟ آیا فاصله بین (دو اندازه گیری) گزارش شده است؟ آیا بیماران در فاصله میان دو اندازه گیری، وضعیت پایداری داشته اند؟ آیا شرایط (نحوه اجرا، محیط، دستورالعملها) در هر دو اندازه گیری یکسان بوده است؟ آیا ضریب همبستگی درون طبقه ای در نمرات پیوسته، محاسبه شده است؟ آیا ضریب کاپا در نمرات اسمی، فاصله ای و دوازدهی محاسبه شده است؟ آیا در نمرات ترتیبی، کاپای وزنی محاسبه شده است؟	
خطای اندازه گیری: آیا حداقل دو روش سنجش وجود دارد؟ آیا بیماران در فاصله ی میان دو اندازه گیری، وضعیت پایداری داشته اند؟ آیا شرایط (نحوه اجرا، محیط، دستورالعملها) در هر دو اندازه گیری یکسان بوده است؟ آیا در نظریه های کلاسیک آزمون، خطای معیار اندازه گیری، حداقل تغییر قابل تشخیص یا محدوده ی توافق، محاسبه شده است؟	

پرسشنامه	آیا در روش طولی، حداقل دو ابزار استفاده شده است؟ آیا زمان فاصله گذاری (بین دو آزمون) گزارش شده است؟ آیا اتفاقاتی که در فاصله زمانی بین دو آزمون روی داده (به طور مثال، مداخله یا دیگر حوادث مرتبط) به طور کامل توضیح داده شده است؟ آیا نسبت بیماران (مثلا بر اساس میزان بهبودی یا وخامت بیماری) تغییر کرده است؟ الزامات طراحی برای آزمون فرضیه؛ زمانی که استاندارد طلایی موجود نباشد: آیا فرضیه ای مبنی بر تغییر نمرات، پیش از جمع آوری داده ها وجود داشته است؟ آیا فرضیه، جهت، وجود ارتباط یا تفاوت میانگین های مورد انتظار در نمرات تغییر یافته ابزار وجود دارد؟ آیا فرضیه میزان مورد انتظار مطلق یا نسبی همبستگی یا تفاوت میانگین ها را در نمرات تغییر یافته ابزار شامل شده است؟ آیا توصیفی کامل از ابزارهای معیار، ارائه شده است؟ الزامات طراحی برای آزمون فرضیه؛ زمانی که استاندارد طلایی موجود باشد: آیا ملاک به کار برده شده برای تغییرات، می تواند به عنوان یک استاندارد طلایی منطقی، محسوب گردد؟ آیا برای اعداد پیوسته، همبستگی بین نمرات تغییر یافته و سطح زیر منحنی عملیاتی دریافت کننده، منحنی راک، محاسبه می شود؟ آیا برای ابزارهای دو ارزشی، حساسیت و ویژگی (تغییر یافته ها در مقابل تغییر نیافته ها) تعیین شده است؟
نتیجه گیری	آیا پراکندگی کل نمرات در نمونه پژوهش، توصیف شده است؟ آیا درصد نمره پاسخ دهنده ای که پایین ترین نمره کل ممکن را داشت، توصیف شده است؟ آیا درصد نمره پاسخ دهنده ای که بالاترین ترین نمره کل ممکن را داشت، توصیف شده است؟ آیا نمرات و تغییر نمرات (به طور مثال میانگین) برای گروه ها و زیر گروه های مرتبط (بیماران یا عموم مردم)، ارائه شده است؟ آیا حداقل تغییر با اهمیت یا حداقل تفاوت با اهمیت، مشخص شده است؟