



**UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS
CENTRO DE INVESTIGACIÓN TRANSDISCIPLINAR EN PSICOLOGÍA**

**Validación de un modelo de evaluación de riesgo psicosocial laboral, en
trabajadores peruanos y mexicanos**

T E S I S

QUE PARA OBTENER EL GRADO DE:

Doctor en Psicología

P R E S E N T A:

César Ayax Merino Soto

Director de tesis

Arturo Juárez García

Comité revisor

Dr. Arturo Juárez García
Dr. Luís Fernando Arias Galicia
Dr. José Manuel Fernández Arata
Dr. Luís Pérez Álvarez
Dr. Pedro Rafael Gil-Monte
Dr. Sergio Alexis Domínguez Lara
Dr. Eduardo Hernández Padilla

Noviembre, 2022

Agradecimientos

A mi esposa Marisol, por su poderosa influencia sobre los pasos que doy (y daré).

A mis hijos Ajax y Soslán, quienes con su sola existencia gatillan mi intención de ser más y mejor.

A mi familia de origen (Godofredo, Rosa, Rosi, Isila), quienes ayudaron a moldear una visión de mundo para construir un camino con limitaciones superables.

A Arturo, para quien las posibilidades de éxito siempre son infinitas ($p < .01^n$).

A una reducida lista de amigos(as) y colegas, quienes contribuyeron aleatoriamente, en un momento determinado, a seguir el camino hasta la presente meta.

Resumen

Dada la aún emergente situación de del desarrollo de modelos de factores psicosociales del trabajo (FPST) en Perú, en contraste con el avance en México de la reciente construcción del modelo de Procesos Psicosociales del Trabajo (PROPSIT), el objetivo del presente estudio fue validar el instrumento derivado del modelo PROPSIT, específicamente de las dimensiones de factores psicosociales, en trabajadores peruanos, y comparar su equivalencia de medición con trabajadores mexicanos. Conducido con un diseño parcialmente mixto, el estudio utilizó predominantemente procedimientos cuantitativos, y complementariamente un procedimiento cualitativo, bajo el marco unificado de validez. Las evidencias de validez de contenido fueron evaluadas en trabajadores peruanos, en tres aspectos, con los siguientes resultados: a) la identificación de la estructura conceptual de los FPST en Perú ($n = 48$), hallándose fuerte convergencia con las dimensiones de factores de riesgo psicosocial pre-establecidas del PROPSIT-FPST, excepto por algunos conceptos emergentes en las dimensiones de recursos favorables, como la recreación en el trabajo; b) la lecturabilidad, claridad y relevancia de todos los ítems del PROPSIT-FPST fue satisfactoria ($n = 54$); y c) la estructura de respuesta y validez con criterios externos ($n = 198$) sugirieron la potencial reducción del escalamiento de respuesta (de 7 a 5 opciones), y las asociaciones fueron predominantemente convergentes con autoeficacia ocupacional, entusiasmo laboral, estrés y síntomas depresivos. Respecto a las evidencias de procesos de respuesta ($n = 359$), se detectaron 54 (14.9 %) participantes con potenciales respuestas descuidadas (esfuerzo insuficiente), y su impacto mediante un análisis de sensibilidad mostró efectos sobre la calidad de los parámetros psicométricos (confiabilidad y tamaño de las cargas factoriales). En las evidencias de estructura interna, a)

la dimensionalidad y relación ítem constructo convergieron en la estructura dimensional esperada, excepto por eliminación de algunos ítems; b) la propiedad convergente y divergente interna fue teóricamente consistente; c) la equivalencia entre trabajadores peruanos (n = 305) y mexicanos (n = 409) fue satisfactoria; y d) la confiabilidad puntual (> .77) y condicional fueron adecuadas. En las evidencias de relación con variables externas de efectos emocionales (experiencia de estrés y síntomas depresivos) y laborales (entusiasmo, accidentes laborales menores y dificultades cognitivas), se obtuvieron asociaciones monotónicas teóricamente convergentes. Se discuten las implicaciones teóricas y de generalización del modelo teórico del PROPSIT, y las limitaciones para la generación de estudios nacionales y comparaciones interculturales.

Palabras clave: factores psicosociales del trabajo, validez, equivalencia de medición, estrés laboral, entusiasmo laboral

Abstract

Given the still emerging situation of the development of models of psychosocial factors at work (FPST) in Peru, in contrast to the progress in Mexico of the recent construction of the Psychosocial Processes at Work (PROPSIT) model, the objective of the present study was to validate the instrument derived from the PROPSIT model, specifically the dimensions of psychosocial factors, in Peruvian workers, and to compare its measurement equivalence with Mexican workers. Conducted with a partially mixed design, the study used predominantly quantitative procedures, and complementarily a qualitative procedure, under the unified validity framework. Evidence of content validity was evaluated in Peruvian workers, in three aspects, with the following results: (a) the identification of the conceptual structure of the FPST in Peru ($n = 48$), finding strong convergence with the pre-established PROPSIT-FPST dimensions of psychosocial risk factors, except for some emerging concepts in the dimensions of favorable resources, such as recreation at work; b) the readability, clarity and relevance of all PROPSIT-FPST items were satisfactory ($n = 54$); and c) the response structure and validity with external criteria ($n = 198$) suggested the potential reduction of response scaling (from 7 to 5 options), and the associations were predominantly convergent with occupational self-efficacy, job enthusiasm, stress and depressive symptoms. Regarding the evidence of response processes ($n = 359$), 54 (14.9 %) participants were detected with potential neglected responses (insufficient effort), and their impact through sensitivity analysis showed effects on the quality of psychometric parameters (reliability and size of factor loadings). In the evidence of internal structure, a) the dimensionality and item-construct relationship converged in the expected dimensional structure, except for the elimination of some items; b) the internal

convergent and divergent property was theoretically consistent; c) the equivalence between Peruvian (n = 305) and Mexican (n = 409) workers was satisfactory; and d) the point (> .77) and conditional reliability were adequate. In the evidence of relationship with external variables of emotional effects (experience of stress and depressive symptoms) and occupational effects (work engagement, minor occupational accidents and cognitive difficulties), theoretically convergent monotonic associations were obtained. The theoretical and generalization implications of the PROPSIT theoretical model and the limitations for the generation of national studies and cross-cultural comparisons are discussed.

Key words: psychosocial work factors, validity, measurement equivalence, work stress, work engagement.

ÍNDICE

Agradecimientos

Resumen

Abstract

1.	Antecedentes o fundamentos teóricos y/o empíricos	6
1.1.	Marco sustantivo.....	7
1.1.1.	Contexto disciplinar y teórico general.....	7
1.1.2.	Contexto teórico general	11
1.1.3.	Desarrollo del PROPSIT (Procesos Psicosociales del Trabajo)	17
1.2.	Marco intercultural	21
	Validación intercultural	23
1.3.	Marco metodológico.....	27
	Validación psicométrica	28
	Estándares consensuados	31
	Equivalencia e invarianza de medición	33
1.4.	Marco contextual	38
2.	Planteamiento.....	42
2.3.1.	Metas.....	50
2.3.2.	Preguntas de investigación	51
2.3.3.	Objetivos.....	52
2.3.3.1.	Objetivos de validez de contenido:.....	52
2.3.3.2.	Objetivo de evidencias de validez de procesos de respuesta.	53
2.3.3.3.	Objetivos de evidencias de validez de estructura interna	54
2.3.3.4.	Objetivo de evidencias de validez con variables	55
3.	Método	56
3.1.	Horizonte teórico-metodológico del investigador.....	56
3.2.	Diseño de investigación	57
3.2.1.	Presunciones	57
3.2.2.	Diseño de la investigación	59
3.3.	Participantes.....	62
3.4.	Instrumentos	64
3.5.	Procedimiento	72
3.5.1.	Enfoque general	73
3.5.2.	Procedimientos en la etapa piloto.....	75
3.5.3.	Procedimientos en la etapa principal	79
3.5.4.	Aspectos éticos.....	84
3.	Resultados	86
3.1.	Objetivo 1a: Identificar la estructura conceptual de los factores psicosociales del trabajo (exposiciones)	88
3.2.	Objetivo 1b: Validez de contenido: Claridad y relevancia de los ítems del PROPSIT	113
3.2.1.	Sección 1: Análisis de la lecturabilidad del contenido de PROPSIT - FPST	116
3.2.2.	Sección 2: Análisis de las competencias del lenguaje de los participantes (literacy)	119
3.2.3.	Sección 3: Análisis del Consenso / acuerdo	125
3.2.4.	Sección 4: Análisis de la claridad de los ítems.....	126
3.2.5.	Sección 4: Análisis de la relevancia del contenido de PROPSIT – FPST	127
3.2.6.	Conclusión general.....	129
3.3.	Objetivo 1c: Validez de contenido: estructura de respuesta de los ítems y validez con criterios externos.....	132
3.4.	Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST	158
3.5.	Objetivo 3a: Identificar la estructura dimensional del PROPSIT.....	178
3.6.	Objetivo 3b: Obtener evidencias de validez convergente y divergente interna del PROPSIT.	205
3.7.	Objetivo 3c: Examinar la equivalencia de medición del PROPSIT	214
3.8.	Objetivo 3d. Examinar la confiabilidad de los puntajes del PROPSIT	235
3.9.	Objetivo 4. Examinar la relación de los constructos del PROPSIT – FPST con otras variables	250
4.	Puntos finales e implicaciones del proyecto	272
5.	Referencias	279
	Apéndice	348

1. Fundamentos teóricos y/o empíricos

El presente proyecto está dentro del marco de la investigación sobre la medición de factores psicosociales del trabajo, específicamente mediante el Cuestionario de Evaluación de Procesos Psicosociales del Trabajo (PROPSIT), un instrumento creado en México para evaluar los factores psicosociales, sus efectos en la salud, y posibles moderadores entre estos factores y efectos. Para esto, el respaldo teórico se distribuyó en bases conceptuales sobre los factores psicosociales del trabajo, y la teoría de la validez. Dado que esta investigación adaptó el PROPSIT para para los trabajadores peruanos, también se desarrollaron contenidos sobre el proceso de adaptación intercultural de medidas, el marco emic/etic para comprender la potencial generalización de los constructos medidos por el PROPSIT, y se realizó una descripción de la situación peruana que requiere de medidas de evaluación de factores psicosociales del trabajo.

La sección introductoria del presente proyecto está organizada en cuatro marcos heurísticos: el sustantivo, el intercultural, el metodológico y el contextual. El marco sustantivo expone los conceptos básicos y resume los modelos teóricos relevantes alrededor de la creación y adaptación del PROPSIT en México; el marco intercultural describe los enfoques que sirven para comprender una investigación de constructos psicosociales en diferentes grupos culturales, y su vínculo con el proceso de validación psicométrica; el marco metodológico desarrolla la conceptualización de la validación psicométrica, la utilidad de estándares o guías para el proceso de validación, y la especificación de la equivalencia de la medición, la cual es un concepto procedimental clave en la presente investigación; finalmente, el marco contextual presenta el desarrollado de los factores

psicosociales en Perú, y su relación con la medición de los mismos. Este último marco sirve como puente para el planteamiento del problema, la justificación y los objetivos o propósitos de la investigación. Posteriormente, la sección del método se organiza con relación al horizonte teórico-metodológico del investigador, una propuesta conceptual inédita y creada en el presente proyecto, y con relación al diseño de investigación, se explican las presunciones principales para realizar el estudio, así como la identificación del diseño propiamente dicho. El desarrollo del método continúa con las características potenciales de participantes, la información relevante sobre los instrumentos, los procedimientos de recolección de datos, el plan de análisis, y las consideraciones éticas del desarrollo del estudio. Finalmente, el documento cierra con una descripción detallada e integrativa de las implicaciones teóricas de los hallazgos obtenidos desde cada objetivo, e implicaciones instrumentales para el desarrollo del modelo PROPSIT y su potencial generalización a otros contextos.

1.1. Marco sustantivo

1.1.1. Contexto disciplinar y teórico general

Contexto general evolutivo

En la etapa adulta del desarrollo humano, el *rol de trabajador* es una de las funciones críticas para la continuidad de la adaptación individual, la cual es construida desde la infancia, y conduce a la preparación de etapas posteriores de desarrollo. El rol de trabajador se define con su adquisición objetiva y con el momento en que ocurre. Este momento puede ser entendido como *estadios de carrera*, donde el crecimiento cronológico de la persona (e.g., edad) está vinculado con diferentes etapas de transición entre una y otra, y con su ajuste personal dinámico (Lindström, 2011). El rol del trabajador puede actuar como

un moderador de la asociación entre los estresores laborales sobre la salud y bienestar, así como dar significancia al estresor y a los recursos de afrontamiento utilizados. Esta relación entre la etapa de carrera, estresores laborales y las respuestas a los mismos, ha sido expuesta principalmente en estudios de países occidentalizados (Cohen, 1991; Hurrell & Lindström, 1992; Lindström & Hurrell, 1992; Dorociak, Rupert, & Zahniser, 2017), en donde la etapa de carrera adquirida por el individuo ha tenido efectos moderadores. Pero su relevancia para contextos como el latinoamericano, aún está sujeta a evaluación empírica, porque los estudios sobre el papel moderador de la etapa de carrera del trabajador aparentemente son escasos o inexistentes.

En la investigación actual, hay evidencias que sugieren que la variabilidad correlacional entre los aspectos del trabajo y los efectos psicosociales en el trabajador pueden explicarse parcialmente por el desarrollo del rol del trabajador (Cohen, 1991; Dorociak et al., 2017). Uno de los efectos psicosociales laborales de mayor trascendencia es la experiencia del estrés en el contexto laboral, la cual ha mantenido una relación persistente con un rango amplio de condiciones laborales físicas y psicosociales (Ganster, & Rosen, 2013).

Epidemiología ocupacional

Una de las disciplinas implicadas en las investigaciones primigenias sobre los efectos en la salud debido a condiciones laborales, es la epidemiología ocupacional, la cual consiste en la aplicación y estudio de enfoques de investigación donde se identifica la distribución y exploración de la relación entre la exposición de condiciones de trabajo y estados de salud (Porta, 2014). Esta disciplina vincula la epidemiología con un proceso de investigación, el cual busca explorar las variables que mantienen o deterioran la salud en ambientes ocupacionales y, por tanto, también apunta hacia la investigación etiológica y vigilancia de

la enfermedad en el trabajo (Guidotti, 2000). Esta área parece puede estar subsumida dentro de la *epidemiología social*, enfocada en la aplicación de la metodología epidemiológica sobre los factores sociales (sean estructuras o procesos) y su relación con la salud y enfermedad (Porta, 2014). Un término más apropiado para describir la implicación de la psicología con la epidemiología sería la *epidemiología psicosocial*, pero parece ser representado como equivalente a *epidemiología social* (Porta, 2014). La comprensión de esta disciplina debe de recurrir a cuerpo de conceptos médicos con referencia a la enfermedad y la estadística, así como su relación a la actividad ocupacional.

Psicología de la salud ocupacional

En un intento de posicionar la psicología en el estudio de la salud en el trabajo y sus correlatos, surgió la psicología de la salud ocupacional (PSO) como una disciplina aplicada y orientada hacia la promoción y protección de la calidad de vida del trabajador, relevante a la dinámica de las exposiciones sociales, psicológicas y organizacionales a la salud del trabajador (Patlán, 2017; Quick, 1999a, 1999b; Schaufeli, 2004). La PSO puede reconocerse como el puente de confluencia de la medicina, la sociología y la psicología, en que las exposiciones sociales en el trabajo, y las respuestas de salud física y mental son enlazadas en los análisis conceptuales y empíricos (Quick, 1999b). Para su consolidación, hubo eventos críticos en Estados Unidos y Europa los cuales influenciaron fuertemente su evolución, tales como como la construcción de instituciones, investigación en base modelos teóricos de alcance mundial, organizaciones profesionales y actividades académicas en el tema (Barling & Griffith, 2003). Un avance actual es la expansión de su cobertura hacia constructos positivos, que influyen consecuentemente en la identificación y promoción del bienestar y rendimiento laboral (Bakker & Rodríguez-Muñoz, 2012).

Aparentemente, psicología de la salud ocupacional (PSO) aún no está bien identificada y consolidada en países hispanos como una disciplina diferenciada y fuertemente independiente a otras ramas de la psicología, sino más bien complementaria a la medicina (e.g., en Cuba: Almirall & Hernández, 2017); o está declarada como una rama emergente desde hace más de una década en países como México (e.g., Juárez-García, 2007) o incluso actualmente emergente en países como Perú (e.g., en Perú: Fernández-Arata, Calderón-De la Cruz, & Navarro-Loli, 2016). En este sentido, su desarrollo sólido aún puede tardar años, más aún en un contexto de conocimientos emergentes. Por ejemplo, en Perú se ha creado un diplomado que se ocupa de los riesgos psicosociales en el trabajo y de la psicología de la salud ocupacional (Universidad Peruana Cayetano Heredia s/f), orientado a la profesionalización de la evaluación de los riesgos psicosociales laborales. Sin embargo, el diplomado está compuesto por un cuerpo docente principalmente de áreas no afines a la psicología (e.g., administradores o ingenieros), y por lo tanto los conocimientos pueden estar altamente parcializados y desvinculados con los avances provenientes de los estudios de la disciplina psicológica. En contraste, en otros países como México, el desarrollo programático y diferenciación con la salud ocupacional médicamente orientada, están mejor reconocidos (Juárez-García, 2007; Patlán, 2017). Finalmente, el desarrollo de la PSO ha tenido un punto de convergencia con la creación de la Red de Investigadores sobre Factores Psicosociales (Juárez-García, García, & Vera, 2010), una reciente plataforma para reducir el crecimiento aislado de la investigación sobre factores psicosociales del trabajo en países hispanos, así como facilitar la comunicación, intercambio y generación de conocimientos sobre el tema.

1.1.2. Contexto teórico general

En esta subsección, se presentan conceptos esenciales que centralizan el grupo de variables sobre el que se desarrollan los objetivos de la presente investigación.

Estrés laboral

En una reciente revisión sistemática realizada sobre 203 artículos anglosajones seleccionados entre 1993 y 2017 (Burman & Goswami, 2018), se halló que por lo menos existen 10 definiciones de estrés laboral. Un análisis no formal del contenido de estas definiciones puede identificar la siguiente comunalidad conceptual entre ellas: el estrés laboral es una experiencia displacentera, que compromete una respuesta multidimensional, cuyo proceso elemental es el desbalance entre la intensidad de la exigencia laboral y las capacidades del trabajador, con el consiguiente decremento de la salud física, psicológica, y el desempeño laboral.

Sobre estas conceptualizaciones del estrés laboral también convergen las propuestas definicionales de entidades influyentes en la gestión de la salud en el trabajo, como el National Institute for Occupational Safety and Health (1999) y la International Labour Organization (2016), que enfatizan la desigual proporción entre las demandas laborales y las respuestas del trabajador. La diferencia entre estas definiciones de estrés laboral frente a la definición del estrés general es que los estresores provienen de contextos donde la persona realiza un trabajo.

Las características conceptuales del estrés laboral llevan a reconocer la existencia de condiciones laborales que generan estrés, esto es, estresores particulares del trabajo que influyen en el desequilibrio exigencias – capacidades. Desde una visión general, la

respuesta inapropiada del trabajador ante el tipo y magnitud de una demanda específica (estresor), es el concepto básico de estrés laboral para comprender sus efectos negativos sobre el trabajador (Schabracq, 2003). En contraste, también hay contextos laborales que promueven la reducción del estrés y elevan la percepción del bienestar y de la salud física, es decir, que colaboran con el trabajador para alcanzar la definición establecida de salud (World Health Organization, 1986). Este tipo de contexto laboral puede identificarse como *trabajo saludable* (World Health Organization, s/f), donde las demandas laborales son apropiadas las capacidades del trabajador. Para precisar el constructo de contexto laboral, el siguiente párrafo lo describe.

Contexto laboral

La situación o contexto laboral, para caracterizarlo mejor, consiste en la organización, el entorno social, contenido de la tarea y su ejecución (Gil-Monte, 2009). Esta definición identifica a la situación laboral como una entidad conceptual multidimensional, interrelacionada en sus componentes, y que actúan sinérgicamente sobre el trabajador. Como se señaló anteriormente, el contexto laboral es una de las fuentes importantes en la vida del adulto, en la que experimenta estresores y estímulos positivos relevantes a su experiencia vital, dentro del continuo malestar-bienestar. Estos estresores y estímulos positivos laborales se han entendido como factores psicosociales del trabajo (FPT); ¿específicamente qué son los FPT? En el siguiente párrafo se aclara el concepto.

Factores psicosociales del trabajo

La definición de *factores psicosociales en el trabajo*, desarrollada por numerosas fuentes, se centra en algunos conceptos centrales, como la situación laboral psicosocial, su

efecto de salud física y psicológica, y la interacción entre ambos. Este último componente conceptual es clave, y ha sido establecido como esencial para comprender el estrés laboral (International Labour Organisation, 1986), y sus efectos sobre el estado de salud del trabajador. La preocupación por los factores psicosociales del trabajo por parte de los organismos internacionales relacionados con la salud y el trabajo, es una respuesta que parece haber aumentado a lo largo de los años, como se pueden reconocer en las publicaciones relevantes (e.g., International Labour Organisation, 1986; Kalimo, El-Batawi, & Cooper, 1987; Houtman, Jettinghoff, & Cedillo, 2007; Leka, Jain, & World Health Organization, 2010; World Health Organization, 2007).

En el contexto del estrés laboral, existieron algunos avances claves que emergieron para comprender y actuar sobre los riesgos de salud del trabajador en el contexto laboral. Primero, la 27ma Asamblea Mundial de la Salud organizada por la Organización Mundial de la Salud (OMS) en 1974, puso en relevancia el efecto causal de los factores psicosociales sobre la salud, lo cual significó no solo definir mejor los constructos asociados a ellos sino también ver el vínculo entre la salud y estos factores externos e internos al sujeto, y que van más allá de las características tangibles del ambiente laboral. Segundo, la comisión mixta de la Organización Internacional del Trabajo y la Organización Mundial de la Salud (1984) avanzó en resaltar la interacción entre las condiciones externas y características de la persona. Esta interacción se comprende mejor en relación con procesos evaluativos del trabajador sobre la situación o estímulo estresante, y es de donde provendría una parte de la varianza explicada del efecto de las condiciones externas sobre los resultados de ajuste o de salud del trabajador.

Si esta interacción es evaluada subjetivamente, desde la percepción del trabajador, entonces hay un ángulo interno, cognitivo, que debe ser tomado en cuenta. Por lo tanto, la interacción entre los factores externos y la capacidad de la persona, se puede explicar parcialmente por la percepción de la capacidad y la oportunidad de control de la persona sobre la situación estresante y la intensidad del estresor (Lazarus & Folkman, 1986), la cual es un componente clave para comprender la activación de las estrategias de afrontamiento, así como la severidad del impacto producido por los estresores (Lazarus & Folkman, 1986).

Tercero, el surgimiento del modelo de *Tensión Laboral o Demanda/Control* (Karasek, 1988), el cual es un modelo predominante en la comprensión de los factores psicosociales del trabajo (Fernandes & Pereira, 2016), tiene como ejes a dos factores de naturaleza psicosocial y continua: las *demandas psicológicas* de las tareas laborales y el *control* que el trabajador puede ejercer sobre las mismas. Este aporte tuvo una seria transformación cuando se agrega el papel del apoyo social (Karasek & Theorell, 1990).

Luego del surgimiento del modelo de *Tensión Laboral o Demanda/Control*, otro modelo que ha mostrado fuerte influencia en la comprensión de los factores psicosociales del trabajo es el de *Desequilibrio Esfuerzo-Recompensa*, enfocado en el concepto de intercambio social (Siegrist & Peter, 2000). Los dos constructos fundamentales de este modelo son el *esfuerzo* y la *recompensa*, entendidos como interacciones de intercambio durante la actividad laboral, entre el trabajador y el ambiente laboral. Desde una visión orientada hacia el tratamiento o la intervención, este modelo y el de Tensión Laboral o Demanda/Control, hay evidencias de su efectividad cuando se les trata complementariamente entre ellos, una situación que eleva la capacidad predictiva de varios resultados de salud estudiados (Fernández-López, Fernández-Hidalgo & Siegrist, 2005;

Ostry, Kelly, Demers, Mustard, & Hertzman, 2003). La implementación conjunta de dos modelos para comprender o intervenir el efecto estresor del trabajo, no es un problema cuando se valora la utilidad de ambos para su integración a estrategias de cambio e intervención ocupacional, así como para explicar el estatus de salud y la presencia de condiciones crónicas de enfermedad (Fernández-López et al., 2005; Ostry et al., 2003).

Los trabajos de revisión sistemática corroboran la consistente acumulación de evidencias de la capacidad predictiva del modelo de Demanda/Control (e.g., Alves, Hökerberg & Faerstein, 2013; van der Doef & Maes 1999) o del modelo Esfuerzo-Recompensa (e.g., Backé, Seidler, Latza, Rossnagel & Schumann, 2012; van Vegchel, de Jonge, Bosma & Schaufeli, 2005), implicados en la comprensión de la respuesta de bienestar (o su ausencia) del trabajador. Este tipo de evidencias confirman el grado de importancia y vigencia de ambos modelos para la investigación.

La capacidad predictiva de estos u otros modelos también pone un reto en la manera de medir sus constructos, especialmente respecto a la cantidad de factores psicosociales que pueden ser contenidos en un instrumento, y su relevancia cultural específica. Estos retos, y especialmente la concerniente a la relevancia cultural, son posibles motivos que parecen haber influido sobre la creación de instrumentos para la medición de factores psicosociales en diferentes países (ver en los párrafos más adelante), que contrastan en los factores psicosociales elegidos por ellos. Unos motivos adicionales pueden ser, primero, de tipo económico como consecuencia del costo para adquirir y renovar la licencia de uso en las editoriales que los publican, y segundo, de los límites metodológicos implicados para su adaptación intercultural. Este último requiere de conocimientos metodológicos avanzados

para enfocarse en los variados aspectos de la validación de una medida en un contexto de comparación intercultural.

Dado que los estresores relevantes en la vida de la persona adulta provienen predominantemente de su situación laboral, los modelos de factores psicosociales se han centrado en los *factores de riesgo psicosocial laboral*.

Factor de riesgo psicosocial laboral

Consiste en el conjunto de características de la situación laboral que conducen a un perjuicio, daño o reacción negativa y desadaptativa del trabajador (Gil-Monte, 2009, 2012). Pero las características de la situación laboral no solo pueden ser negativas para el bienestar del trabajador, sino también positivas cuando favorecen la actividad laboral y el desarrollo saludable del trabajador (Gil-Monte, 2009). Este efecto positivo o negativo de un factor psicosocial sobre la salud del trabajador es lo que lo define en alguna dirección de ambas valencias o polos (es decir, negativo o positivo). Esta distinción de los factores psicosociales del trabajo en relación a su efecto tiene algunas implicaciones: primero indica que éstos pueden ser operacionalizados en algún punto del continuo de efectos negativos y positivos; por ejemplo, las demandas laborales pueden ser definidas como excesivas demandas (para enfatizar su efecto negativo), demandas apropiadas (para enfatizar su efecto positivo), o simplemente, demandas laborales (para enfatizar un punto neutral pero que abarca ambos efectos). Segundo, que se requiere un proceso moderador de la relación del factor psicosocial y la respuesta (o efecto) consecuente, que está establecido teóricamente en los modelos más importantes del tema.

Las implicancias del efecto negativo o positivo sobre el trabajador proveniente de los factores psicosociales, se pueden entender desde al menos tres aspectos (Gil-Monte, López-Vílchez, Llorca-Rubio, & Sánchez-Piernas, 2016): la salud ocupacional (es decir, las respuestas adaptativas o desadaptativas del trabajador), el impacto económico (que indica la pérdida de beneficios propios y para la organización laboral), y la prevención (esto es, la orientación hacia la seguridad y reducción de las condiciones de riesgo por parte de la organización laboral, y el compromiso del trabajador para cumplirlos). Junto a estos aspectos, también converge el incremento del rendimiento laboral y calidad de la productividad y servicios. Aunque los factores psicosociales del trabajo pueden conducir a respuestas negativas o positivas del trabajador (Gil-Monte, 2009, 2012) y ambos deben recibir interés en la investigación y en los servicios de intervención, la atención generalmente se ha puesto en los factores asociados con el estrés desadaptativo (International Labour Organisation, 1986; Lu, Nakata, Park, & Swanson, 2014).

1.1.3. Desarrollo del modelo de Procesos Psicosociales del Trabajo (PROPSIT)

El presente proyecto está influenciado por dos aspectos: primero, el desarrollo de un modelo evaluativo de factores, procesos y efectos psicosociales laborales, denominado actualmente PROPSIT (*Procesos Psicosociales del Trabajo*, Juárez, 2017; Juárez & Camacho, 2011), y que contiene una propuesta directa de estimar el grado de interacción de las demandas laborales y recursos personales/laborales. Este componente interactivo es la principal fuerza para comprender sus efectos psicosociales sobre el trabajador (Bakker & Demerouti, 2007; Demerouti, Bakker, Nachreiner & Schaufeli, 2001). Segundo, la experiencia de 21 años de investigación (que incluye hallazgos empíricos y desarrollos metodológicos) sobre estos aspectos psicosociales del trabajo en trabajadores mexicanos

(Juárez, 2017). De acuerdo a este modelo, el conjunto de estos componentes (demandas y recursos en el contexto del trabajo) son instrumentales para comprender los factores relevantes para los problemas de salud o psicológicos derivados del contexto laboral, y para planificar las estrategias de validación de una nueva medida de factores psicosociales laborales. El modelo de evaluación PROPSIT, de acuerdo a los autores (Juárez, 2015, 2017; Juárez & Camacho, 2011) está fuertemente influenciado por el Modelo Demandas-Recursos (Bakker & Demerouti, 2007; Demerouti, Bakker, Nachreiner & Schaufeli, 2001). Este modelo tiene características que resaltan frente a otros modelos, como la visión heurística de los estresores y sus efectos aplicados al contexto laboral, la inclusión de resultados y procesos de tipo positivo y negativo entre sus componentes, la vinculación entre áreas de aplicación de salud ocupacional y recursos humanos, y la incorporación de los recursos individuales del trabajador (Demerouti, & Bakker, 2011; Schaufeli & Taris, 2014). Debido que el modelo Demandas-Recursos (D-R) es de amplia cobertura conceptual, los constructos de otros modelos también pueden ser subsumidos dentro de él. Por ejemplo, los componentes del modelo Esfuerzo-Recompensa (E-R), y Demanda-Control (D-C) no se contradicen con el modelo D-R, y pueden integrarse como factores psicosociales específicos. En este camino, el PROPSIT también incorpora a los factores de los modelos E-R y D-C debido a su potencial cobertura etic (e.g., universal).

Las cuestiones clave para el desarrollo del PROPSIT fueron varios aspectos teórico-conceptuales, y prácticos. Primero, fue la estructuración del concepto de factores psicosociales, derivado de la revisión de los conceptos existentes de los factores psicosociales, y desarrollos conceptuales en México (Juárez, & Camacho, 2011); de este modo, para lograr una cómoda transición de la teoría a la práctica de evaluación e intervención, los factores psicosociales laboral son definidas como condiciones de trabajo

que interactúan con el trabajador, y que mediante mecanismos patogénicos y salutogénicos asociados al estrés influyen en la mantención de su salud y/o de enfermedades (Juárez-García, 2017, pp. 28; Juárez-García, 2018, pág., 96). Otro aspecto teórico fue el reconocimiento del proceso de interacción como la pieza fundamental que vincula los factores psicosociales y las respuestas adaptativas o desadaptativas en el contexto laboral. La interacción sugiere la actuación dinámica, no estática, de las características del ambiente psicosocial y las características individuales del trabajador; esto significa que la interdependencia de ambos, y en que para la comprensión completa del efecto (estrés o bienestar), necesariamente debe involucrarse el análisis o evaluación de esta relación (e.g., la interacción). La interacción como principio clave es un componente teórico y operacional de modelos como el de Karasek & Teorell (1990), que describe la interacción de las demandas laborales y el grado de autonomía del trabajador; o del modelo de Siegrist (Siegrist, & Peter, 2000), que describe la interacción de las recompensas del ambiente laboral y el esfuerzo del trabajador. Finalmente, el modelo de Bakker y Demerouti (2007) también incluye un proceso interaccional entre las demandas laborales y los recursos del trabajador.

Otro aspecto teórico fue la identificación de factores psicosociales salutogénicos y patogénicos del contexto laboral. Los primeros también pueden ser identificados como los aspectos favorables o saludables en la medida que promueven el bienestar personal, la productividad y la satisfacción del trabajador; los segundos (factores patógenos) están vinculados con lo que se pueden considerar la exposición a riesgos, y que usualmente causan el deterioro del bienestar, la productividad y satisfacción (Juárez, en prensa; Juárez & Camacho, 2011). Ambos tipos de factores (patogénicos y salutogénicos) se pueden entender como extensiones ubicadas en un continuo correspondiente con la expresión

positiva y negativa del ambiente laboral. Por ejemplo, una expresión patogénica puede ser la violencia entre pares que ocurre en el trabajo, habitualmente conocida como mobbing, mientras en la orientación opuesta de este factor estaría el apoyo social. Es improbable que coexistan simultáneamente en el mismo ambiente, o asociados al mismo agente causal. Finalmente, la aparente generalización de factores de riesgo en varios contextos culturales, derivado de la información empírica acumulada a lo largo de los años, ha llevado a reconocer que éstos aparecen en muchas formas de actividad laboral, como la autonomía del trabajador, las recompensas, y las excesivas demandas. Una presentación gráfica del modelo aparece en la Figura 1.

Por otro lado, algunos puntos prácticos también fueron identificados para iniciar el desarrollo del PROPSIT (Juárez-García, 2018; Juárez-García & Flores-Jiménez, 2020). Primero, el restringido contenido muestreado en los instrumentos existentes sobre las variables causales (determinantes del estrés o bienestar), las variables de respuesta (efectos teóricamente vinculados a las variables causales), e interacciones entre estas variables. Para el contexto mexicano, el análisis de esta situación en la medición de los factores psicosociales, y otros instrumentos derivados en otros contextos, ha verificado un número variable de factores, operacionalizados en cuestionarios de auto-reporte. Esta revisión (Juárez-García, 2018; Juárez-García & Flores-Jiménez, 2020) ha conducido a identificar dos componentes de un modelo basado en integrar, en cada componente, el continuo patogénico-salutogénico, esto es, un *sistema de factores psicosociales del trabajo* de valencia positiva (recursos favorables) y valencia negativa (riesgos psicosociales), como fuente de los efectos psicosociales del trabajo, también de valencia positiva (bienestar y entusiasmo laboral) y valencia negativa (malestar psicológico y agotamiento).

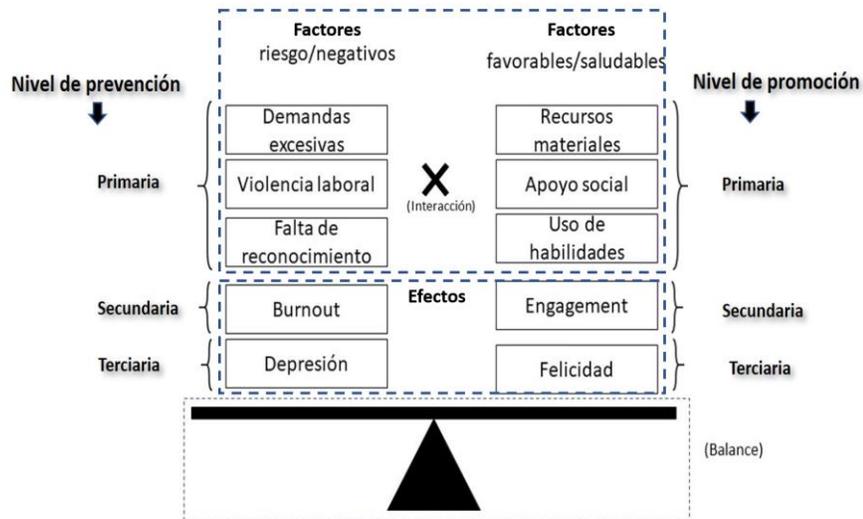


Figura 1

Balance de factores psicosociales positivos (saludables) y negativos (de riesgo). Adaptada de: Juárez-García & Flores-Jiménez (2020)

1.2. Marco intercultural

En el tipo de investigación que aquí se propone (validación intercultural de una medida), existen dos enfoques habitualmente involucrados: *emic* y *etic*. La utilización de estos enfoques está influenciada por los trabajos de Berry (1969, 1989), quien los sistematizó para su aplicabilidad en los estudios interculturales en psicología. El enfoque *etic* conduce a identificar y enfatizar conceptos y conductas de aplicación universal y generalizables entre diferentes locaciones donde pueden emerger (Niblo & Jackson, 2004), y desde el cual se pretende describir y comprender la experiencia de la persona con una visión de comunalidad de atributos. Esto es, que la experiencia del individuo puede ser comprendida desde el observador externo, mediante significados asumidos como comunes y generalizables, y que pueden ser transferidos desde un contexto cultural a otro (Berry, 1969, 1989). Por otro lado, el enfoque *emic* se sustenta en la variabilidad de las conductas y conceptos delimitados por el contexto cultural específico en que ocurren, de tal modo que

es más apropiado con mejor exactitud comprender la experiencia de la persona, desde él mismo, pero no desde un observador externo. En esta visión, la aplicación del significado de un atributo, y la construcción de un atributo, no podrían generalizarse a otros contextos. Ambos enfoques son fuertemente influyentes para la metodología de investigación concerniente a la validación de instrumentos, porque un diseño de investigación puede ser conducido en la línea de tres diferentes métodos: *etic impuesto*, *emic paralelo* o *etic derivado* (Berry, 1969, 1989).

El primero corresponde con un diseño en que el instrumento elegido, creado en una cultura diferente, se aplica con la presunción que los constructos medidos y la situación de aplicación del instrumento se mantienen similares en sus significados. Esta es la perspectiva frecuentemente utilizada en la investigación que usa una medida proveniente de una cultura diferente, pero es una de las prácticas más cuestionables (Niblo & Jackson, 2004). En contraste con el anterior, el enfoque *emic paralelo* dirige la creación de medidas y constructos *nativos*, dentro de cada cultura, y explícitamente sensibles a una determinada cultural. En este enfoque, un constructo es creado para representar la experiencia subjetiva de la persona en su propia cultura, sin la pretensión de generalizarlo a otras culturas. Finalmente, la integración eficaz para desarrollar constructos que sean sensibles culturalmente, pero también que logren capturar la comunalidad del constructo en múltiples contextos, corresponde con el enfoque *etic derivado*. Procedimentalmente, este enfoque puede expresarse en secuencias de métodos que generalmente combinan el enfoque cuantitativo y cualitativo (Bala, Chalil, & Gupta, 2012; Berry, 1989, 1990; Niblo & Jackson, 2004). Operativamente, su aplicación requiere explorar los constructos desde un enfoque *emic*, para construir conocimientos y su naturaleza asociativa, que conducen finalmente a relevar una estructura de conceptos *nativos*. En una siguiente fase de análisis,

los conceptos nativos son evaluados desde un enfoque etic, para identificar patrones que los unen conceptualmente con constructos correspondientes a otro contexto cultural.

De este modo, debido a la aplicación secuencial de los ambos enfoques (i.e., emic y etic), el enfoque etic derivado es potencialmente útil para proporcionar validez intercultural a los constructos medidos por el PROPSIT, que corresponde a lo apropiado de la interpretación de sus puntajes. La validez, enfocada de esta manera, permite que el contenido de las evaluaciones tome en cuenta las diferencias culturales de los grupos a quienes se destina el instrumento (Solano-Flores & Milbourn, 2016), y por lo tanto que sea sensible a diferencias individuales en el constructo más allá de las variaciones culturales.

En este aspecto, las lecciones posiblemente más aceptadas durante décadas en la validación de medidas de uso intercultural en psicología, parecen provenir de la investigación de la personalidad (e.g., Fetvadjev, & Van de Vijver, 2017), donde se han establecido la robustez de atributos de personalidad usando varias estrategias de evaluación de la equivalencia. Pero en estos estudios, la principal contribución metodológica es la conjunción de los enfoques emic y etic, para identificar atributos psicológicos universales y específicos a una cultura.

Validación intercultural

Para la comparación de grupos culturalmente diferentes, en ciencias sociales usualmente se incluyen atributos, actitudes u opiniones, y sus instrumentos de captura de respuesta son centrales para definir el constructo medido. Pero la centralidad fundamental para la comparabilidad es la medida misma, es decir, que la herramienta sea funcionalmente comparable, o similar, respecto a lo medido. Este proceso es conocido como *adaptación cultural* (Beaton, Bombardier, Guillemin, & Ferraz, 2000), referido como

la adaptación del contenido y del lenguaje (mediante la traducción), de una medida desarrollada y utilizada en una *cultura-fuente* (*source*), para su uso en una *cultura-objetivo* (*target*). Con relación a los *estándares de medición* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), la validez intercultural no es una faceta independiente de validez, como lo es la validez de la estructura interna, o la relación convergente/divergente entre constructos. Desde una visión heurística, puede ser enmarcada como una evidencia de la validez de constructo (van Raaij, 1978). De acuerdo con la interpretación prevalente y actual de la validez, esto significa que la interpretación del puntaje, y sus características conceptuales se mantienen constantes entre las culturas comparadas, de modo que las variables latentes son entidades que explican similarmente el comportamiento proveniente de diferentes culturas.

Una búsqueda casual de la literatura centrada en la adaptación intercultural de medidas parece revelar que no existe un apropiado balance entre los objetivos de investigación y la metodología involucrada para cumplirlos, de tal modo que algunos escenarios de adaptación parecen descuidados, y otros intensamente enfatizados. Por ejemplo, la documentación prevalente sobre la validación o adaptación intercultural de medidas psicosociales se refiere a: 1) los procesos idóneos de traducción de un idioma a otro, especialmente del inglés al español), 2) el empleo de sofisticados análisis cuantitativos aplicados a variables latentes (e.g., modelamiento de ecuaciones estructurales y teoría de respuesta al ítem), y 3) la aplicación de las metodologías propuestas en unos escenarios de adaptación, con descuido de otros escenarios (e.g., un instrumento en idioma español en creado en un contexto, pero aplicado en otro contexto también de habla española). De particular importancia para el presente estudio es el último punto, porque corresponde a un

escenario que tiende a ser menos elaborado en la literatura de adaptación de medidas psicosociales (McDowell, Courtney, Edwards, & Shortridge-Baggett, 2005; Vallejo-Medina et al., 2017) y, por lo tanto, de una menor extensión de guías, revisiones o argumentaciones sobre las limitaciones y procedimientos implicados.

Para describir mejor esto, existen cuatro posibles escenarios en que el proceso de adaptación de una medida puede requerir diferentes modificaciones, y consecuentemente, diferentes pasos en la implementación de la metodología (Beaton et al., 2000; Guillemin, Bombardier, & Beaton, 1993). Éstos se refieren al uso del instrumento en 1) en inmigrantes en la cultura-fuente, 2) en otro país, pero el mismo idioma, 3) en inmigrantes de la misma cultura-fuente, pero diferente idioma, y 4) en otro país y otro idioma (Beaton et al., 2000; Guillemin et al., 1993).

Para el presente estudio, el escenario relevante es el que requiere usar un instrumento (en este contexto, el PROPSIT) creado en México (cultura-fuente) en otro país (Perú, cultura-objetivo), pero en que ambos mantienen el mismo idioma oficial (e.g., castellano). Los procesos involucrados para adaptar el PROPSIT en este escenario no requieren principalmente la traducción total ni parcial del lenguaje, sino la adaptación cultural del contenido, presumiendo que se pueden esperar diferencias importantes en la significancia cultural de los constructos (Beaton et al., 2000; Guillemin et al., 1993). Sin embargo, no hay una literatura relevante identificable para este escenario, que se encuentre desarrollada con amplitud y profundidad para servir de guía al investigador. La existente guía (e.g., Vallejo-Medina et al., 2017) es aún emergente en su aplicabilidad y consenso.

Recientemente (Vallejo-Medina et al., 2017) se ha estructurado una guía sugerida para la adaptación de medidas en el escenario descrito (e.g., diferentes contextos, pero el mismo idioma), y se caracteriza por ser una útil referencia orientada hacia la práctica. Sin

embargo, las sugerencias específicas en varios puntos parecen requerir la misma fuerza para su racionalidad, así como modificaciones para expandir su protocolo propuesto. Para el presente proyecto, se elaboró un análisis crítico de las sugerencias realizadas por Vallejo-Medina et al., (2017), con la finalidad de fortalecer una nueva propuesta de adaptar medidas entre contextos culturales que mantienen el mismo idioma. En la Figura 2 aparecen los argumentos de las posibles modificaciones, que potencialmente son la base causal para proponer una nueva guía de adaptación relevante al presente estudio.

Número de guía y contenido (Vallejo-Medina et al. (2017))	Racionalidad de modificación
1. Elegir la versión validada apropiadamente	Ninguna
2. Guías estandarizadas del proceso	Ninguna
3. Elección de expertos	a) La percepción de los participantes-objetivo también es una fuente principal de información b) Puede existir discrepancia entre expertos y participantes, y puede ser subestimada con solo elegir jueces expertos. c) Con apropiada instrucción en el procedimiento, los expertos pueden ser elegidos en base a su experiencia profesional.
4. Criterios de evaluación del fraseo del contenido (entendimiento, extrañeza, potencial re-fraseo)	d) Los criterios son altamente dependientes, polarmente opuestos, y causalmente relacionados. e) Pueden ser evaluados mediante un solo criterio, por ejemplo, claridad.
5. Fraseo alternativo	f) El fraseo alternativo puede generar expresiones locales, que maximizan la comprensión en el grupo target, pero reducen la generalización básica del fraseo. g) El fraseo alternativo puede ser insustancial y motivado por las preferencias idiosincrásicas del experto.
6. Compilación de la evaluación	h) Categorizar en base al número de jueces que sugieren modificaciones subestimada la magnitud de la modificación propuesta. i) La inclusión de participantes requiere añadir el peso de sus juicios comparada con los juicios de los jueces expertos.
7. Decisión de cambio	j) La decisión también requiere valorar la magnitud (“tamaño del efecto”) de las sugerencias de modificación propuestas. k) La inclusión de participantes requiere añadir el peso de sus juicios comparada con los juicios de los jueces expertos.
8. Evaluación de equivalencia	l) La percepción de los participantes-objetivo también es una fuente principal de información
9. Evaluación de propiedades de contenido	m) La claridad y el entendimiento parecen solaparse fuertemente, y los jueces no podrían hacer distinciones entre un criterio u otro. n) El tamaño del instrumento puede influenciar en la indisponibilidad de los jueces para resolver cinco criterios para cada ítem, y esto a su vez en producir calificaciones espurias.

Figura 2
Racionalidad de potenciales limitaciones en el protocolo de adaptación (Vallejo-Medina et al., 2007)

1.3. Marco metodológico

El avance de la globalización del mercado del trabajo, las empresas y servicios, parece ser un elemento común para impulsar la necesidad de evaluar la aplicabilidad intercultural de constructos relevantes a los FPS del estrés en el trabajo (Sanchez, Spector & Coper, 2006). Por este motivo, existe un aparente impulso para el desarrollo de estudios interculturales y generar conocimientos interculturalmente transferibles entre los contextos, aún con la certeza de potenciales diferencias culturales-específicas en cada contexto de aplicación. Esto es particularmente importante en el tema del estrés en el trabajo, porque existen claras implicancias para la salud y para la economía (personal y global), independientemente del contexto cultural.

Lo anterior es crítico porque en la evaluación de los factores psicosociales laborales, hay desafíos y dificultades originados en la operacionalización de variables que son esencialmente subjetivas, es decir, procesos o estados psicológicos sin representación tangible ni directamente medibles, pero identificados desde la percepción de la persona mediante instrumentos cuantitativos. Junto al problema de evaluar procesos no directamente observables, se suma otro problema: la metodología de recolección de datos mediante medidas de auto-informe requiere la percepción de la persona sobre el estado o proceso en cuestión (Sánchez et al., 2006). Estas dificultades se refieren específicamente a varios puntos (Rick & Briner, 2000): a) la falta de precisión para identificar con claridad la fuente o estresor, b) la poca inmediatez de sus efectos, c) el efecto variable (de positivo a negativo, o de negativo a positivo), y d) la evaluación cognitiva del trabajador respecto al peligro (*hazards*) y el daño (*harm*) producido. Quizá la controversia más severa es determinar el vínculo entre el grado de daño psicosocial (i.e., enfermedad) y el factor

psicosocial presente o aparentemente causante, e integrarlo con herramientas de evaluación accesibles (Tabanelli et al., 2008) y claros en su uso (Potocka, 2012).

Validación psicométrica

En la descripción y evaluación de una medida sobre el comportamiento psicológico, los términos comúnmente utilizados de validez, confiabilidad o propiedades psicométricas no significan lo mismo. Dentro del rango de definiciones conceptuales académicas en psicología, parecen existir pocas y detalladas descripciones sobre el significado de *propiedades psicométricas*. Sin embargo, este detalle aparece en Osterling (2008), quien lo define como alguna característica técnica y esencial de una medida, expresada comúnmente en información numérica, y que provee lo necesario para interpretar sus puntajes. Llevado a la práctica de investigación, la definición operacional de una propiedad empírica difícilmente puede centrarse en un solo número representativo, debido que varias propiedades específicas usualmente están subsumidas en una propiedad genérica (Osterling, 2008).

Para desarrollar el presente proyecto dentro de una metodología coherente, se decidió utilizar el marco *unificado de la validez*, referido a un enfoque donde los aspectos de contenido, estructura interna, de asociación con otras variables, y consecuencias de un instrumento se tratan de manera inter-relacionada para probar empíricamente inferencias sobre la interpretación y uso de los puntajes obtenidos de un instrumento (Messick, 1994). Lo anterior contrasta con el modelo tradicional de la validez, enfocada en el test *per se* (Kane & Bridgeman, 2017), y donde las propiedades psicométricas pueden interpretarse estáticamente en el instrumento y se recurren a estrategias cuantitativas para evaluar tres

componentes de validez: el contenido, la naturaleza conceptual constructo, y su asociación con criterios conductuales o de desempeño. En cambio, el enfoque unificado actual enfatiza varias facetas que guían la obtención de múltiples evidencias empíricas, todas enfocadas unitariamente para validar a) la interpretación del constructo, y b) sus fines científicos o aplicados (Kane & Bridgeman, 2017; Messick, 1994). En la perspectiva unificada, la *validez del constructo* subsume a las otras fuentes de validez, y por lo tanto las evidencias sobre el uso e interpretación de los puntajes aseguran la medición de constructo con el instrumento analizado. Incluso la estimación de la confiabilidad de la medición está integrada como evidencia de validez de la estructura interna.

La observación casual de algunas publicaciones de validez en revistas de alto impacto puede reconocer la influencia de este enfoque en la comunicación de los procedimientos de validez psicométrica, en que las palabras clave *evidencias*, *interpretación*, *uso* y *puntajes* son sucedáneos aceptados para referirse a la *validez del instrumento*, o *confiabilidad del instrumento*. Por ejemplo, es aceptado actualmente que la confiabilidad es una propiedad de los puntajes, y no del propio test, y las evidencias de validez se aplican a las inferencias de los puntajes, no al mismo instrumento (AERA et al., 2014). El presente estudio aplicará esta recomendación en la expresión de los resultados y procedimientos.

En la operacionalización de una visión integrada de procedimientos de validez, una de las más influyentes fue de Loevinger (1957), quien propuso que la validez de constructo se basa en tres estadios: validez *sustancial*, *estructural* y *externa*. Posteriormente, desde los trabajos de Messick (1994, 1995) se estructuran los modos de evidencia de validez en cinco aspectos. Por otro lado, influenciados principalmente por este último trabajo, y en coherencia con el marco integrativo de la metodología mixta de investigación, también se

ha planteado un *modelo de meta-validación* (Onwuegbuzie, Bustamante & Nelson, 2010; Onwuegbuzie, Daniel, & Collins, 2009), el cual forma parte de un marco de validez más amplio, y estructurado para resolver varias áreas de la validez de una investigación desde la aplicación de los procedimientos de la metodología cuantitativa y cualitativa (Onwuegbuzie & Johnson, 2006). Este marco diferencia las fuentes de evidencia en dos grandes áreas de obtención de evidencias: evidencias con *respaldo lógico*, y con *respaldo empírico*. La evidencia basada en el respaldo lógico se obtiene mediante los procedimientos habituales de *validez de contenido*, pero en este modelo meta-analítico se diferencia en *validez de face*, *validez del ítem* y *validez del muestreo del contenido*. Por otro lado, la validez con respaldo empírico contiene procedimientos para obtener evidencias de *validez relacionada con criterios*, como el uso de criterio concurrente (*validez concurrente*) o un criterio predictivo (*validez predictiva*). Finalmente, la *validez relacionada con el constructo*, se orienta a las evidencias obtenidas mediante procedimientos de validez sustantiva, estructural, comparativa, de generalización y consecucional.

Un enfoque de validación claro ayuda a decrementar la práctica de uno de los problemas aún relevantes en la investigación, que es la *inducción de la validez* (Merino-Soto, & Angulo-Ramos, 2020; Merino-Soto, & Calderón-De la Cruz, 2018), un concepto que describe la tendencia a utilizar las evidencias de validez obtenidas de estudios antecedentes o realizados en otros contextos, para respaldar afirmaciones de validez en los datos actuales. Este concepto es una extensión de la *inducción de la confiabilidad*, en que se realizan el mismo tipo de afirmaciones, pero respecto a la confiabilidad de los puntajes (Vacha-Haase, 1998). En ambas situaciones de inducción, ocurre la omisión de evidencias críticas para comprender las limitaciones y alcance de los instrumentos. Otro problema que se incrementa sin el contexto de guías claras es la traducción de los ítems de un

instrumento, que pueden ser una fuerte importante para generar respuestas no válidas respecto al contenido evaluado, lo cual es especialmente cierto cuando el idioma de origen es diferente al idioma-objetivo (Sanchez et al., 2006). Afortunadamente, existen guías normalizadas para estructurar las evidencias de validez de medidas; y algunas de estas serán introducidas en el siguiente párrafo.

Estándares consensuados

Normas prototípicas. En base a la estructura conceptual sobre los componentes de la validación de un constructo, actualmente hay guías y estándares aceptados internacionalmente para la adaptación o evaluación de tests (e.g., International Test Commission, 2016; Lindley, Bartram, & Kennedy, 2008). Estas guías o estándares son herramientas consensuadas por equipos de investigadores internacionales en psicometría y evaluación, y sirve de apoyo para que el investigador pueda estructurar los pasos básicos en la adaptación de tests. Este consenso internacional integra la opinión experta y los reportes de buenas prácticas de adaptación de medidas, y es públicamente accesible.

Para lograr los objetivos del presente proyecto, también se tomó como respaldo a las directrices del Internacional Test Commission, ITC (2016), debido a su influencia internacional para estructurar la obtención de evidencias de validez, y su sensibilidad a la tendencia globalizadora del uso de los tests (Muñiz, Elosua & Hambleton, 2013). Los procedimientos lógico-rationales y empíricos utilizados en el presente estudio también respondieron a las directrices establecidas por la ITC, en una estructura correspondiente con la Figura 3.

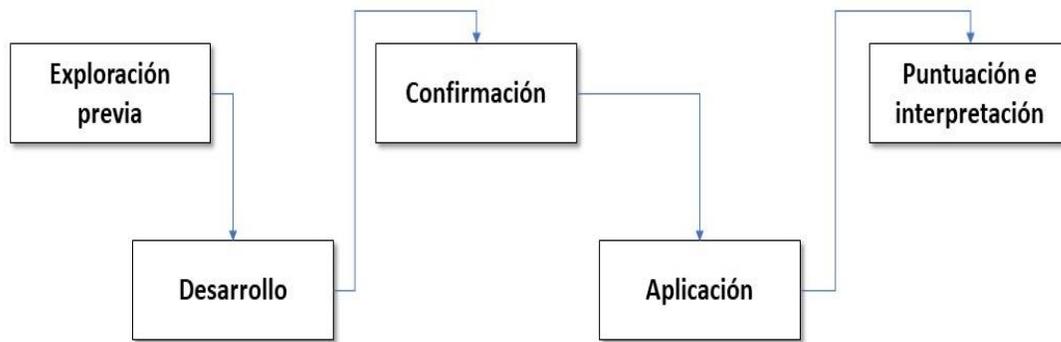


Figura 3

Directrices secuenciales y generales de la International Test Commission (ITC) para la adaptación de medidas psicológicas

Los pasos mostrados en la Figura 3 representan una visión consensuada, heurística, y general para efectuar un proceso de adaptación de un instrumento de factores psicosociales del trabajo. Efectivamente, estas guías estandarizadas, como la propuesta por el ITC, representan también la revisión del estado del arte de las guías existentes, y su unificación es un avance para la tecnología en la construcción y adaptación de medidas en psicología y ciencias afines, pero también aplicables a medidas de factores psicosociales del trabajo.

Esta literatura metodológica aplicada a la adaptación de instrumentos se ha focalizado principalmente en un solo escenario, esto es, la adaptación de una medida construida en un lenguaje diferente (Vallejo-Medina et al., 2017). La situación prototípica para esto es la elección de una medida desarrollada para la población angloparlante, y sobre la cual se aplica en proceso de adaptación hacia una población hispanohablante. Para esta situación, existen apenas pocas referencias de la literatura (Vallejo-Medina et al., 2017) y por lo tanto hay un espacio vacío de conocimiento consensuado para ejecutarlo. Lo que

frecuentemente ocurre es generalizar los principios creados para adaptaciones *inter-idioma* (e.g., del inglés al castellano) hasta esta situación *intra-idioma* (e.g., del castellano hablado en México al castellano hablado en Perú). En la sección anterior se hizo un análisis crítico de este problema, mientras que también se propuso unas iniciativas de modificación (ver Figura 2) del protocolo elaborado por Vallejo-Medina et al. (2017).

Equivalencia e invarianza de medición

La conceptualización de la invarianza de medición está principalmente enmarcada en la metodología estadística y matemática, y es una herramienta metodológica subsumida en la investigación de la *equivalencia de la medición* para verificar que una medición produce resultados estadísticamente equivalentes o comparables, en dos o más grupos (Allen & Walsh, 2000; Fetvadjev, & Van de Vijver, 2017). La conceptualización de la equivalencia de una medida está respalda por taxonomías de los tipos de equivalencia propuestos en la literatura. Por ejemplo, se pueden describir cuatro formas de equivalencia de una medida (Allen & Walsh, 2000): de la traducción, métrica, conceptual y función. En una vía similar, y asociado a la metodología estadística de la invarianza de medición, se puede diferenciar la equivalencia de constructo, métrica y escalar (Fetvadjev, & Van de Vijver, 2017; He, & Van de Vijver, en prensa); otros modelos incluyen la equivalencia de muestreo (Buil, de Chernatony, & Martínez, 2012), o más parsimoniosamente, de constructo y de medición (Poortinga, 1989).

Amenazas a la equivalencia. La equivalencia entre-grupos (e.g., entre trabajadores mexicanos y peruanos que responden al PROPSIT) de una medida puede debilitarse debido a varias fuentes de sesgos. De acuerdo con la más reciente diferenciación de las amenazas a la equivalencia intercultural de una medida (Byrne, 2016), se pueden reconocer al menos

tres grandes clasificaciones: sesgo de *constructo*, de *método* y del *contenido de los ítems*. El primero (sesgo de constructo) se refiere a la significancia diferencial del constructo entre los grupos culturales, y generalmente es el problema más difícil de resolver en la adaptación de medidas (Byrne, 2016). Este se expresa de dos maneras: a) cuando los contenidos de la medida insuficientemente representan la variabilidad del constructo, es decir, el dominio conductual del constructo está subrepresentado en el grupo cultural donde se lo aplica; y b) cuando el contenido cultural es diferencialmente relevante, esto es, los objetos culturales referidos en una medida son desconocidos o diferentemente entendidos.

El segundo sesgo referido es el de método, asociado a la *variabilidad de las muestras* de participantes (*sesgo de muestreo*), que posiblemente es uno de los más desafiantes procesos en la investigación comparativa en general, y en la investigación intercultural en particular. En la investigación intercultural es inusual lograr un diseño con muestreo aleatorio (He, & Van de Vijver, en prensa), y es más probable obtener muestras de participantes desde un muestreo no probabilístico, específicamente de conveniencia (He, & Van de Vijver, en prensa). Otro muestreo más controlado de selección muestral incluye el muestreo sistemático (Boehnke, Lietz, Schreier, & Wilhelm, 2011), en que se maximiza la representatividad y la similaridad de los grupos (Boehnke et al., 2011; van Raaij, 1978). Por otro lado, la presentación o formato del contenido del instrumento (*sesgo instrumental*), y los procedimientos de aplicación (*sesgo de administración*) son otros aspectos del sesgo del método potencialmente presentes. La tercera y última amenaza a la equivalencia proviene del sesgo del ítem, referido como el contenido específico que produce interpretaciones diferentes entre los grupos, aun cuando el constructo sea similar en ambos.

Todas las formas de sesgo descritas anteriormente confluyen en el significado esencial de amenaza a la equivalencia de una medida, es decir, que una medida tiene el

problema potencial que no elicitara respuestas similares debido a aspectos ajenos al propio constructo.

Invarianza de medición. La evaluación de la equivalencia requiere una prueba estadística que formalice los pasos intermedios de este proceso, y es habitual que se adopte un conjunto de pasos cuantitativos jerárquicamente aplicados de análisis psicométricos (Byrne, 2016). Procedimentalmente, esto tiende a referirse al estudio de la *invarianza de medición*, como un método para verificar la retención de las propiedades métricas de una medida en diferentes grupos de personas, en un diseño de comparación de grupos donde se presume que la pertenencia a un grupo pueden ser fuente de variabilidad de estas propiedades (Byrne, 2016; Joreskog, 1971). El término es usualmente aplicado en el estudio de variables latentes (Fetvadjev, & Van de Vijver, 2017) dentro del marco del modelamiento de ecuaciones estructurales (*structural equation modeling*, o abreviadamente, SEM; Joreskog, 1971), y por lo mismo, la metodología estándar para cumplimentar la propiedad de invarianza es SEM. Este método permite obtener los parámetros estadísticos requeridos (Cheung, & Rensvold, 2002; Mokkink et al., 2010; Prince, 2008), como las cargas factoriales y los interceptos de las variables observables (ítems), y medias, varianzas y covarianzas de las variables latentes.

Existen otros enfoques basados en estimaciones de variable latente no lineales, como el estudio del funcionamiento diferencial del ítem (*differential ítem functioning*, o DIF) dentro de enfoques paramétrico (como la Teoría de Respuesta al Ítem, o regresión logística), o no paramétricos, como la aplicación de tablas de contingencia (Teresi, 2006). La aplicación conjunta de alguna de estas metodologías puede ser vía de rigurosidad del estudio de la invarianza, y puede ser preferido en lugar de la implementación de una sola metodología (Teresi, 2006).

Los modelos matemáticos para probar la invarianza inician con una formulación que describe la relación entre los indicadores manifiestos (variables observables o ítems) y sus correspondientes variables latentes. Para esto, la estructura de medición de las variables observadas (Liao, 2002; Steenkamp & Baumgartner, 1998; Brown, 2015), la conceptualización básica es que el puntaje obtenido en un ítem es descrito por un modelo de regresión lineal, donde la respuesta observable x_i al ítem i está compuesta por un intercepto τ_i , un coeficiente beta (λ_{ij}) que relaciona el constructo latente ξ_j a la respuesta x_i , y finalmente el término de error aleatorio δ_i :

$$x_i = \tau_i + \lambda_{ij}\xi_j + \delta_i \tag{1.1}$$

La representación conceptual de esta estructura de medición se presenta en la Figura 4, donde el óvalo representa la variable latente o constructo, los cuadriláteros representan las variables observadas, indicadores o ítems.

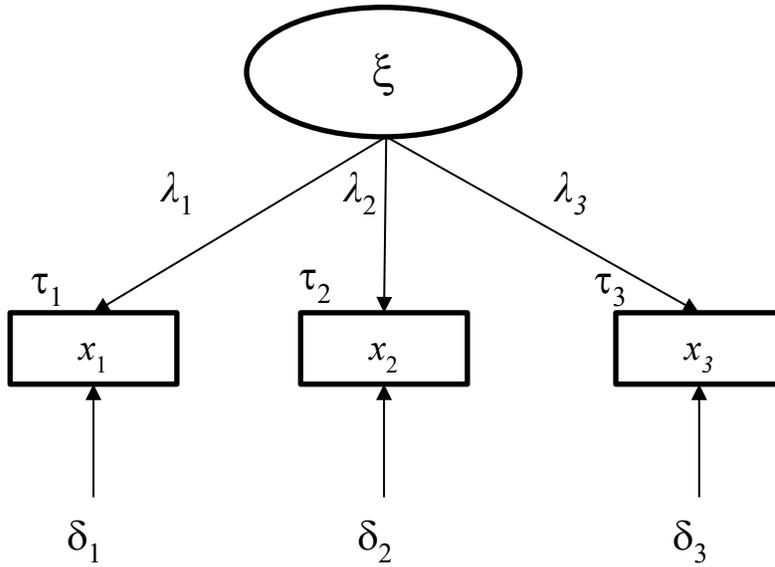


Figura 4

Representación del factor latente (ξ), indicadores observables (x), cargas factoriales (λ), interceptos (τ) y error de medición (δ)

Esta representación en el contexto de la invarianza de medición, incluye la identificación del grupo g (g), para indicar la fuente de variabilidad de los grupos comparados:

$$x_i^g = \tau_i^g + \lambda_{ij}^g \xi_i^g + \delta_i^g \quad (1.2)$$

La anterior notación describe a un único indicador observable, pero la representación multivariada completa de un conjunto de ítems se expresa en notación algebraica matricial. Este conjunto relacionado se expresa mediante la matriz de covarianza

de los ítems (Σ) y sobre ella la influencia causal de la matriz de cargas factoriales (Λ), la matriz de varianzas y covarianzas entre las variables latentes o constructos estimados (Φ) y la matriz diagonal de las varianzas de error de los ítems (Θ_{δ}); para propósitos del cálculo, se añade la traspuesta de Λ (Λ'):

$$\Sigma^g = \Lambda^g \Phi^g \Lambda'^g + \Theta_{\delta}^g \quad (1.3)$$

Cada uno de los anteriores componentes son los objetivos particulares del estudio de la invarianza de medición. De este modo, la invarianza configuracional es la primera prueba que se evalúa, donde se examina la igualdad del número de variables latentes y los demás parámetros (i.e., cargas factoriales, interceptos y residuales) se estiman sin restricciones de igualdad entre los grupos A y B. Segundo, la invarianza métrica entre el grupo A y B se establece con la identidad de las matrices de cargas factoriales entre los grupos, estableciéndose como hipótesis nula: $H_o : \Lambda^A = \Lambda^B$. El siguiente nivel de invarianza (escalar) aplica la restricción de igualdad del vector de interceptos, es decir, $H_o : \tau^A = \tau^B$. Finalmente, el último nivel de invarianza (varianza de error o residuales) requiere la restricción de igualdad de la matriz de varianzas y covarianzas de error:

$$H_o : \Theta^A = \Theta^B.$$

1.4. Marco contextual

Para conocer el contexto de la cultura-objetivo en que se aplicará el instrumento (PROPSIT), una pregunta fundamental debe plantearse: ¿Cómo se encuentra el desarrollo y los avances de la investigación sobre los factores psicosociales del trabajo en Perú? Hasta la

fecha, el desarrollo de la investigación científica teórica o aplicada en Perú, respecto al tópico, puede ser identificada en un estado de desarrollo emergente, en relación con estudios psicométricos, como no psicométricos (Merino-Soto & Juárez-García, 2019). Desde dos estudios de revisión sistemática de las publicaciones peruanas relevantes a los FPS (Merino-Soto & Juárez-García, 2019), se puede compendiar la situación peruana como una línea base de estado actual que no solo comprende un estrecho rango de tiempo, sino un número de años que acumulativamente han contribuido a las características del avance peruano en FSP. Estos dos estudios también forman parte de los objetivos de la presente tesis.

En pocos años se ha avanzado con la validación psicométrica de medidas relacionadas con los FPS (Calderón-de la Cruz, Domínguez, & Arroyo-Rodríguez, 2018; Calderón-De la Cruz, Merino-Soto, Juárez-García, Jiménez-Clavijo, 2018; Fernández & Merino, 2014; Fernández, Juárez, & Merino, 2015; Fernández, & Merino, 2016; Flores, Fernández, Merino, Guimet, & Juárez, 2015; Manzanares-Medina, Merino-Soto, & Fernández-Arata, 2016; Merino-Soto, Juárez-García, Altamirano, & Velarde, 2018; Pando-Moreno, Varillas, Aranda-Beltrán, & Elizalde-Núñez, 2016; Suárez, 2013); estos trabajos han sido realizados por tres tipos de alianza de trabajo: autores peruanos, autores peruanos y no peruanos, y autores no peruanos.

Estas validaciones cubren un estrecho rango de evidencias de validez, porque 1) se limitan principalmente a evaluar la estructura interna, y específicamente la dimensionalidad, 2) obtener estimaciones de un solo tipo de confiabilidad (consistencia interna), 3) fueron desarrolladas mayoritariamente en trabajadores de Lima Metropolitana, 4) las decisiones sobre los cambios en los ítems son principalmente empíricas, sin justificación teórica o respaldada por revisión de expertos o de los cambios, 5) se

desconoce si se documentaron y controlaron tendencias de respuesta irrelevantes al constructo, y especialmente referidos a la deseabilidad social, 6) no se verificó alguna equivalencia de las propiedades psicométricas entre grupos, y 7) la interpretación de los resultados generaliza inapropiadamente el alcance de la evidencia de validez específica investigada, hacia otros tipos de evidencia, una situación que se conoce como inducción de la validez (Calderón-De la Cruz et al., 2018).

Solo recientemente, se han reportado trabajos en que un conjunto de evidencias de validez, incluida la estructura interna, han proporcionado resultados más informativos (Calderón-De la Cruz et al., 2018; Merino-Soto et al., 2018), pero aun manteniendo un solo tipo de estimación de la confiabilidad, y con pequeños tamaños muestrales que requieren verificación de su replicabilidad.

En resumen, del marco contextual peruano se puede concluir aun no parece posible concluir sobre si los instrumentos validados en Perú son defendibles para usarlos en la práctica evaluativa de los FPS, dado que el reducido número de evidencias publicadas no permite la acumulación de la información psicométrica relevante. Si se comparan con el tipo de evidencias de validez mínimas requeridas que se pueden plantear para evaluar a una medida como satisfactoria para determinado uso, es posible afirmar que aún falta un largo camino para avanzar. Sin embargo, estos hallazgos plantean que los esfuerzos de validación van incrementándose en número y complejidad, y facilitan su uso para otros grupos de investigación (Merino, 2019a).

Por otro lado, la literatura peruana psicológica, publicada en el mismo país alrededor de los factores psicosociales del trabajo, está aún en la infancia, dado el muy pausado avance de las líneas de investigación institucionales relevante a la psicología del trabajo (Merino-Soto & Juárez-García, 2019). Específicamente, los avances se caracterizan

por ser esporádicos, con poca continuidad en un rango de tiempo, con debilidades metodológicas y psicométricas, numéricamente pequeña, dispersa en investigadores esporádicos, y centradas en la población trabajadora de Lima Metropolitana (Merino-Soto & Juárez-García, 2019). Una aparente consecuencia de esto es que el interés se mantenga bajo o se decremente en términos de la generación de problemas y líneas de investigación sobre los factores psicosociales, que no exista información rigurosamente obtenida para comprender el tema, y que las decisiones de intervención recurran plausiblemente a la experiencia personal de los consultores, la intuición, y la conveniencia (Merino-Soto & Juárez-García, 2019). En un aparente esfuerzo sistemático por introducir directamente el tema, recientemente se han comunicado opiniones bien argumentadas para actualizar el tema de los factores psicosociales (e.g., Fernández-Arata & Calderón-de la Cruz, 2017; Fernández-Arata, Calderón-de la Cruz, & Navarro-Loli, 2016), y que crean el contexto para atraer el interés entre los investigadores, y orientar con mejor argumentación teórica la investigación.

Un marco general legal actualmente existe en Perú, referida directamente a la salud ocupacional e identificada como el Reglamento de la Ley Nro 29783 sobre la seguridad y salud en el Trabajo (Ministerio de Trabajo y Promoción del Empleo, 2012). En referencia a los aspectos psicosociales, se indica el registro del monitoreo de agentes psicosociales (Artículo 33), la definición de exposición a riesgos psicosociales (Artículo 103), la notificación de situaciones de riesgo para vida psicológica (Artículo 110), y tres menciones en el glosario de términos. Por otro lado, aparentemente, la salud ocupacional como es entendida en Perú como una extensión de los servicios prestados por las especialidades de ciencias de salud e ingeniería, como médicos del trabajo, enfermeros ocupacionales y el

higienista industrial (Ramírez, 2012) no tiene clara la participación del psicólogo en estos servicios.

En la fecha de publicación del Manual de Salud Ocupacional en Perú (Ministerio de Salud, 2005), la información objetiva sobre los accidentes de trabajo y enfermedades ocupacionales no parecía existir (Ministerio de Salud, 2005, pag, 7), y la información al respecto requiere de incorporar avances sobre las investigaciones recientes en trabajadores peruanos, sobre factores psicosociales en general, y sobre burnout particularmente. Aun cuando la investigación peruana sobre el burnout está entre las más productivas en Latinoamérica (Hernández, Cabrera-Orozco, Carranza, Mamani-Benito, & Turpo, 2021), se requiere información científica sobre la aplicabilidad de modelos de factores psicosociales tomando en cuenta la variabilidad cultural peruana, así como más estudios de la calidad métricas de las medidas destinadas para evaluar estos factores.

2. Planteamiento

2.1. Limitaciones actuales

Existe un número identificable de medidas que en conjunto o independientemente abordan la medición de las características de los factores psicosociales en el trabajo, las consecuencias psicológicas de estos factores, y/o las características individuales. Para propósitos didácticos, éstas pueden ser clasificadas de acuerdo con si son *medidas únicas* o una *batería de medidas* (Mercer, 1991). Cualquiera de ellas puede ser *unidimensionales* (e.g, Barraza, 2011) o *multidimensionales* (e.g., el *Maslach Burnout Inventory – General Survey*, MBI-GS; Maslach et al., 1996; *Burnout Measure*; Pines & Aronson, 1988).

En el caso de las medidas unidimensionales, la especificidad de su contenido es una de sus fortalezas, porque el muestreo del contenido del constructo puede ser más amplio en

las facetas de un solo constructo. Desde otro ángulo, esto representa una limitación porque el usuario requiere más de una medida para capturar otros constructos relevantes al objetivo de evaluación. Junto a esta limitación, la elección de otros constructos puede estar influenciada por preferencias personales, la existencia y conocimiento de sus propiedades psicométricas o la accesibilidad de estas. Otra limitación en este enfoque es que, en el contexto aplicado o científico, las medidas pueden ser inconsistentemente elegidas por los evaluadores, lo cual tiene eco sobre la cantidad de evidencias de validez que vinculan a los instrumentos elegidos. Por ejemplo, para evaluar la respuesta a los factores psicosociales del trabajo, y un atributo individual, un evaluador puede elegir el MBI-GS y *Wong-Law Emotional Intelligence Scale* (WLEIS; Wong & Law, 2002) como medidas de burnout e inteligencia emocional respectivamente, y construir su paquete de evaluación con ellas, pero otro evaluador puede utilizar el *Cuestionario de Evaluación del Síndrome de Quemarse en el Trabajo* (CESQT; Gil-Monte, 2005) y el *Schutte Self Report Emotional Intelligence Test* (SSEIT; Schutte et al., 1998) para los mismos propósitos. Aunque estos instrumentos se denominan en relación con el constructo medido, el contenido muestreado de estas medidas no es necesariamente el mismo, la precisión puede ser diferente (respecto a la confiabilidad), y las correlaciones entre ellas raramente son muy altas ($r \geq .90$) o altas ($r \geq .80$), aun cuando sus constructos tengan la misma nominación. Como consecuencia de lo anterior, los resultados obtenidos no pueden ser comparables, y los usuarios pueden incrementar su confusión para interpretar los puntajes.

Por otro lado, las baterías de instrumentos pueden tomar en cuenta varios constructos unidimensionales y multidimensionales en un solo paquete de evaluación, sin embargo, un análisis lógico de esto sugiere una limitación importante: pueden existir tantas baterías de evaluación como equipos de evaluadores existen, y estas variaciones pueden

estar impulsadas por preferencias personales, conocimientos de las evidencias de validez de los instrumentos, o la intuición personal. Una elección sin claridad en los criterios puede llevar a realizar modificaciones de contenido, especialmente cuando provienen de otro contexto cultural. De acuerdo con el grado de modificación percibida, los puntajes obtenidos de estos instrumentos modificados no son necesariamente comparables con sus estudios previos, y las evidencias previas de validez tampoco se mantendrán apropiadas.

Otras potenciales limitaciones de crear baterías de instrumentos previamente construidos se relacionan con las diferencias sustanciales en el número de ítems, de tal modo que la representación de sus constructos puede ser no balanceada en cada área medida. Por otro lado, los cuestionarios elegidos para integrar baterías son valoradas como de cobertura universal, y que abordan constructos de alto orden o de amplio contenido (Jacukowicz, & Wezyk, 2017); sin embargo, el contexto cultural de su elaboración no puede ser evitado, es razonable pensar que se construyó dentro de una visión *emic*, es decir, con contenidos no necesariamente no generalizables a otros contextos.

Otra de las limitaciones inherentes es utilizar cuestionarios largos, cuya extensión en relación con el número de ítems y al tiempo requerido para responderlos, es un problema. Estos problemas limitan la evaluación de factores psicosociales del trabajo, y no son menos importantes que el respaldo de su validez y precisión (Rosário, Fonseca, Nienhaus, & Torres, 2016), debido a su influencia sobre la generación de sesgos de respuesta asociados a su extensión, por ejemplo, cansancio o esfuerzo insuficiente. La revisión sistemática sobre los instrumentos de factores psicosociales realizada en diferentes estudios (Charria, Sarsosa & Arenas, 2011; Magnavita, 2008; Potocka, 2012; Tabanelli et al., 2008) tiende a converger en que hay factores comunes que son evaluados, como los constructos de control y carga de trabajo; pero existen problemas adicionales a los listados

en los párrafos anteriores. Por ejemplo, se puede reconocer la influencia predominante de las medidas creadas en el contexto angloparlante, y desde las cuales algunas veces son derivadas versiones hispanas de las mismas. Aunque esto no parecería ser un problema si se consigue la replicación de sus evidencias de validez, éstas son medidas asumidas de constructos universales pero que pueden no contener factores con relevancia cultural-local. Instrumentos con contenidos contextualmente relevantes pueden representar significados más próximos a las características de la población, la cultura y otros aspectos idiosincrásicos de la cultura en que la medida fue creada.

Finalmente, los instrumentos basados en el auto-reporte del riesgo psicosocial y sus efectos han predominado en la investigación sobre los factores psicosociales del trabajo sobre otro tipo de evaluaciones (Charria et al., 2011; Conway, 2009; Tabanelli et al., 2008). Aunque hay ventajas económicas de este tipo de instrumentos para los estudios de campo y el análisis de datos (Conway, 2009), parece no ser habitual probar si las respuestas a estas medidas han sido afectadas por procesos irrelevantes a sus constructos medidos, como estilos de respuesta, deseabilidad social, aquiescencia, esfuerzo insuficiente, u otros patrones identificados por la literatura (Fisher, Fontaine, van de Vijver & van Hemert, 2009; Harzing, 2006; Van Dijk et al., 2009). Otra fuente de variabilidad irrelevante al constructo ocurre cuando el contenido está orientado positiva o negativamente (Solíz, 2015). Excepto el reciente estudio en habla hispana (Fernández-Arata & Merino-Soto, 2016), en que se exploró la relación entre el MBI-GS (Maslach, Jackson & Leiter, 1996) con la deseabilidad social, no parecen existir otras investigaciones asociadas a este tema, al menos en la literatura hispana. Por lo tanto, aun no es claro si la variabilidad de la respuesta a estos cuestionarios está parcialmente influenciada por estos potenciales problemas, tanto en el contexto profesional y aplicado, como en la investigación.

Junto a las anteriores limitaciones metodológicas, las diferencias de género también han sido una fuente de efectos diferenciales de los factores de riesgo psicosocial sobre los problemas de salud y psicológicos (Mommersteeg, Widdershoven & Kop, 2017). Posiblemente estos efectos están asociados a la vulnerabilidad individual y/o a la distribución de roles, pero se requiere claridad e iniciativas de investigación intercultural orientadas hacia la evaluación de la equivalencia de medición intra-cultural, es decir, dentro de la misma población. Esto también se aplica a cualquier comparación de grupos porque la invarianza de las propiedades métricas en los grupos de interés es un prerrequisito para el análisis comparativo.

2.2. Justificación

Aparentemente, una medida única (en contraste a una batería de medidas) y multidimensional puede ser la opción de medición más viable y comprensiva para abordar varios aspectos de los factores psicosociales laborales, porque una medida así construida es un conjunto unitario pero independiente de contenidos que respondan a necesidades locales, y que posiblemente tiene potencial para contribuir la generación de la teoría subyacente a esta medida. La identificación de este tipo de instrumento no debe sugerir que ésta es la única fuente de obtención de información para evaluar la intensidad de los factores psicosociales del trabajo. Más bien, un instrumento así sirve como operacionalización multidimensional y compacta de factores psicosociales, y como un primer escaneo respaldado por un cuerpo teórico consistente e internamente coherente desde su construcción. En una situación contraria, una batería de instrumentos armada con varias medidas creadas en diferentes contextos puede sugerir el problema que *“el todo no es igual a la suma de sus partes”*.

En este sentido, una opción de medición útil es la que corresponde con los instrumentos únicos multidimensionales, y que sean sensibles a las demandas del contexto laboral local. Algunos de estos han sido desarrollados en Italia (*Multidimensional Organisational Health Questionnaire*, MOHQ; Avallone & Pamplomatas, 2005), en Polonia (*Psychosocial Working Conditions*, PWC; Widerszal-Bazyl & Cieslak 2000), en Dinamarca con el *Copenhagen Psychosocial Questionnaire* (COPSOQ; Kristensen, Hannerz, Høgh, & Borg, 2005), y *QPS Nordic – General Nordic Questionnaire* (QPS NORDIC; Lindstrom et al., 2002). En España, con dos instrumentos representativos: la batería de la *Unidad de Investigación Psicosocial de la Conducta Organizacional*, UNIPSICO (Gil-Monte, 2016a; 2016b) y el *Instrumento para la Prevención de Riesgos Psicosociales* (ISTAS, Moncada, Llorens & Kristensen, 2004), el cual es una adaptación del COPSOQ. En el Reino Unido con el *HSE Management Standards Indicator Tool* (HSE-MSIT; Health & Safety Executive, s/f; Edwards & Webster, 2012), el *Enquête Québécoise sur des Conditions de Travail et la Santé et la Sécurité au Travail* (EQCOTESST; Boulianne & Fortin, 2006) en Canadá, y el recientemente desarrollado *OrgFit* (Jimenez & Dunkl, 2017) en Austria. El avance en los enfoques de evaluación ha llevado también a valorar y crear medidas con contenido específico hacia una profesión, por ejemplo, en músicos (Jacukowicz & Wezyk, 2017), y profesores (Unda, Uribe, Jurado, García, Tovalín, & Juárez, 2016).

Todas las medidas anteriores pueden servir de prototipos para elaboración de medidas de FPST en otros países, respecto al modelo conceptual o las variables específicas dentro de cada modelo. ¿Cuál de estas medidas puede servir mejor para representar los FPST en otros países, por ejemplo, Perú o México? Esta no es una cuestión sencilla de responder, pero puede asentarse que la convergencia del idioma, de la región geográfica, las

características de la fuerza laboral, entre otras, pueden ser criterios relevantes. En una vía aplicada, a) los problemas de traducción y captura del significado son decrementados entre medidas del mismo idioma, b) medidas generadas en Norteamérica o Europa pueden referirse a características locales y no orientadas hacia otros contextos, y c) aun con creciente incremento de la globalización, se espera que la fuerza laboral y su contexto psicosocial sean cualitativamente distintos. A estos ejemplos de diferencias, se añade el estatus de desarrollo del instrumento. Específicamente, medidas desarrolladas desde hace años han establecido los puntos fuertes de la teoría y su aplicación en sus contextos de desarrollos, pero en modelos teóricos en desarrollo, sus constructos componentes y los vínculos entre ellas son flexibles debido a su condición de modelo emergente o en construcción. En este sentido, también sus instrumentos de medición están en desarrollo, y ambos (i.e., teoría e instrumentación) son oportunidades para probar su generalización en otros contextos, y nutrirse de investigaciones que los pongan a prueba.

En países como México y Perú, el reto inicial de hacer evaluaciones de factores psicosociales es contar primero con un modelo generado deductivamente, pero con relevancia cultural. De este modo, se requiere crear una medida que sea localmente relevante (Juárez, 2017), pero también potencialmente generalizable, de un modelo de evaluación de procesos psicosociales del trabajo diseñada desde la necesidad de implementar una nueva metodología estandarizada de administración y uso instrumental con evidencias locales e internacionales de validez (Magnavita, 2008). Esto último también puede significar evaluar la significancia etic y emic de los constructos elegidos, para proyectar su uso en contextos culturales diferentes; por ejemplo, un instrumento creado en México, pero con proyección de uso en Perú.

En este sentido, el modelo evaluativo del PROPSIT actualmente está respaldado por componentes empíricos y teóricos. Empíricamente, hay estudios exploratorios en trabajadores mexicanos (Juárez, 2005, 2017), que identificaron correlaciones entre exposiciones y efectos psicológicos. Teóricamente, hay elaboraciones deductivas de los modelos teóricos internacionales más predictivos y aplicables interculturalmente (e.g., Bakker & Demerouti, 2007; Demerouti et al., 2001) al contexto latinoamericano (Juárez & Camacho, 2011). Junto a los dos respaldos señalados (teórico y empírico), existe un respaldo adicional, de tipo pragmático, para la creación de un proyecto de adaptación psicométrica que sea sustentable a largo plazo. Este proviene de la colaboración internacional antecedente, y en que se ha logrado generar algunos estudios empíricos colaborativos (Fernández, Juárez & Merino, 2015; Flores, Fernández, Merino, Guimet & Juárez, 2015; Flores-Jiménez, Merino-Soto, Camacho-Ávila, Juárez-García & Placencia-Reyes, 2015; Juárez, Vera, Merino, Gómez, Feldman & Hernández, 2014; Merino-Soto, Calderón-de la Cruz, Gil-Montes & Juárez, 2021; Merino-Soto, Juárez-García & Angulo-Ramos, 2017), y en donde han predominado los estudios de validación psicométrica.

La validez de los instrumentos es un importante mediador para hacer inferencias precisas y apropiadas sobre los constructos de interés, en relación con criterios psicológicos importantes. Incluso, un conjunto de buenas propiedades psicométricas ayuda a resaltar la replicabilidad y precisión del efecto de los estresores sobre la salud respecto a otros estudios (Rosário, et al., 2016), en la medida que los puntajes tienen menor error de medición. Por otro lado, hay implicaciones para la investigación y para la práctica derivadas del presente estudio. Por ejemplo, las implicancias para la práctica consisten en que el PROPSIT contribuiría a diseñar un sistema de intervención con base empírica, activando un sistema de monitoreo o evaluación en diferentes tiempos; también, puede

diferenciar niveles de intervención (e.g., primaria, secundaria y terciaria) para distribuir racionalmente las acciones interventivas. Por último, las implicaciones para la investigación van en la vía de su vinculación teórica con el modelo de Demandas-Recursos, y la construcción a largo plazo de un modelo teórico singularizado para la población de trabajadores mexicanos y peruanos.

2.3. Propósito de la investigación

2.3.1. Metas

En esta sección, se presentan las metas del estudio, de las cuales emergen los objetivos específicos y sus preguntas de investigación asociadas.

La meta principal del estudio es evaluar la validez del modelo de evaluación de los factores psicosociales del trabajo, basado en la propuesta conceptual deducida de la acumulación de evidencias empíricas y conceptuales del trabajo de Juárez (2005, 2015, 2017; Juárez & Camacho, 2011). Esta propuesta o modelo evaluativo se operacionaliza mediante la Encuesta de Procesos Psicosociales del Trabajo (PROPSIT), y se recurrirá a la participación de trabajadores de México y Perú para obtener sus respuestas al PROPSIT y evaluar su validez de medición. Este propósito es multifacético y fundamentado específicamente en la validar la interpretación y uso del PROPSIT, lo que implicará evaluar comparativamente la relevancia, claridad e impacto de su contenido, su dimensionalidad, la invarianza de los parámetros psicométricos obtenidos, su relación nomológica con constructos convergentes y divergentes, su relación con criterios de funcionamiento laboral, de salud psicológica y física; y la precisión de los resultados.

La segunda meta es contribuir con el desarrollo teórico del PROPSIT para la población mexicana y peruana de trabajadores, examinando la covariación de sus

constructos, y con indicadores laborales que indiquen el bienestar y funcionamiento laboral. La meta final y tercera, consecuente de las anteriores metas, es de tipo metodológico, porque para la consecución de los objetivos específicos (ver más adelante), se aplicarán por lo menos dos enfoques analíticos asociados a modelos estadísticos diferentes. La finalidad de esta estrategia de remover la dependencia de la interpretación de los resultados a un solo método de análisis, y consecuentemente disminuir el potencial sesgo producido por el método aplicado.

2.3.2. Preguntas de investigación

La pregunta principal de estudio es: ¿las propiedades psicométricas del PROPSIT permiten una interpretación equivalente de sus puntajes entre trabajadores peruanos y mexicanos? Dentro de esta pregunta, se construyeron preguntas específicas correspondientes con el tipo de propiedad que se investigará. En este sentido, se elaboraron las siguientes preguntas, organizadas en los principales tramos del estudio:

1. Para explorar e identificar las propiedades del contenido del PROPSIT que se orienten la obtención de evidencias de validez de contenido:
 - a. ¿El contenido de los ítems del PROPSIT es culturalmente relevante, claro, estructuralmente apropiado y válido con criterios externos, para trabajadores mexicanos y peruanos?

Para evaluar la evidencia de validez respecto al proceso de respuesta involucrada en el PROPSIT:

2. ¿Cuál es el grado de sesgos de respuesta a los ítems del PROPSIT, en trabajadores peruanos y mexicanos?

Para obtener evidencia de la estructura interna:

3. ¿La estructura de dimensiones subyacentes del PROSIT será corroborada en la muestra de trabajadores mexicanos y peruanos?
4. ¿Cuál es la magnitud de convergencia y divergencia de las relaciones internas de las dimensiones del PROPSIT?
5. ¿Los constructos medidos por el PROPSIT son psicométricamente equivalentes en trabajadores mexicanos y peruanos?

Para investigar el grado de precisión de los puntajes, se planteó:

6. ¿Cuál es la magnitud de la confiabilidad y precisión de los puntajes del PROPSIT, en trabajadores mexicanos y peruanos?

Para la evidencia basada en relación con otros constructos:

7. ¿Los constructos medidos por el PROPSIT mantendrán asociaciones teóricamente consistentes con otros constructos?

2.3.3. Objetivos

De acuerdo a las preguntas planteadas anteriormente, se declaran los siguientes objetivos de la presente investigación, y para cada una, la breve racionalidad de su elaboración. En la Figura 5 se presenta un resumen que relaciona los objetivos, y las evidencias de validez seleccionadas en este estudio.

2.3.3.1. Objetivos de validez de contenido:

Objetivo 1a: Identificar la estructura conceptual de los factores psicosociales del trabajo (exposiciones). De acuerdo a la comprensión etic derivada de constructos que son evaluados interculturalmente, la identificación de constructos generalizables requiere primero valorar sus características emic y abstraer sus contenidos que son razonablemente

comunes o étic. Por lo tanto, antes de imponer una estructura conceptual (e.g., proveniente del PROPSIT - FPST) en un contexto diferente de donde fue creado, primero se evaluarán los componentes conceptuales específicos y luego los componentes conceptuales generalizables.

Objetivo 1b: Examinar la claridad y relevancia de los ítems del PROPSIT - FPST. La evaluación habitual de la validez de contenido es obtener por lo menos dos evidencias de la adecuación del contenido con la teoría y propósitos de una medida: la claridad de los ítems, y su relevancia o representatividad en relación con su constructo esperado. Este objetivo es útil incluso para garantizar el proceso étic en la conducción del estudio.

Objetivo 1c: Examinar la estructura de respuesta de los ítems y su validez con criterios externos. La evaluación de la validez de contenido se hizo examinando la asociación de cada ítem con constructos externos al modelo del PROPSIT – FPST. Este objetivo ayuda a definir la validez de los ítems con su asociación esperada con variables externas.

2.3.3.2. Objetivo de evidencias de validez de procesos de respuesta.

Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST. Los sesgos de respuesta no parecen haber recibido amplia atención en la investigación internacional sobre la evaluación de FPT, y especialmente en la creación de instrumentos para esta finalidad. Esto posiblemente ocurre porque la ausencia de sesgos de respuesta es una presunción de la calidad de las respuestas a estas evaluaciones, pero se requiere verificación empírica. Adicionalmente a este argumento, la presente investigación propone que

identificar los patrones irrelevantes de respuesta puede servir parcialmente como evidencia de validez basada en los procesos de respuesta al contenido del PROPSIT – FPST.

2.3.3.3. Objetivos de evidencias de validez de estructura interna

Objetivo 3a: Identificar la estructura dimensional del PROPSIT – FPST. La estructura dimensional corresponde con la identificación de las dimensiones latentes y su relación isomórfica con los ítems que los representan y con los puntajes sumarios obtenidos (e.g., sumatoria de ítems). Lograr una estructura identificable y clara es elemental para definir el algoritmo de la creación de los puntajes observados. También, la estructura dimensional es una de las evidencias estadísticas fundamentales para continuar con otras evidencias de validez, porque éstas requieren definir la estructura latente.

Objetivo 3b: Obtener evidencias de validez convergente y divergente interna del PROPSIT. La fuerza con que una dimensión latente del modelo PROPSIT es diferenciable de otra dimensión dentro del mismo modelo, es esencial para la replicabilidad del constructo, así como para su distinción conceptual. Aunque habitualmente la identificación de la estructura dimensional incorpora las evidencias internas de convergencia y divergencia entre sus constructos, en este estudio se examinan por separado en este objetivo, para proporcionar detallada información sobre estos resultados.

Objetivo 3c: Examinar la equivalencia de medición del PROPSIT – FPST. Este objetivo enfoca el grado comparabilidad de las propiedades psicométricas entre trabajadores mexicanos y peruanos, para garantizar que las posibles diferencias (o ausencia de ellas) de originan en diferencias en el nivel de los constructos, y no en diferencias de calibración del PROPSIT – FPST que enmascara la variabilidad entre-grupos.

Objetivo 3d. Examinar la confiabilidad de los puntajes del PROPSIT – FPST.

Los puntajes observables pueden derivarse luego de definir la estructura dimensional latente, pero los puntajes observables continúan influenciados por la variabilidad aleatoria. Esta variabilidad aleatoria corresponde con el error de medición, el grado en que este error ocurre alrededor del puntaje observado, es importante para fines prácticos. Debido que esta variabilidad de error puede no mantenerse constante entre grupos, o dentro de los niveles del puntaje, este objetivo también analizará la confiabilidad condicional a los grupos (peruanos vs mexicanos) y a los niveles del puntaje (desde el mínimo puntaje hasta el máximo puntaje posible).

2.3.3.4. Objetivo de evidencias de validez: asociación con otras variables

Objetivo 4. Examinar la relación de los constructos del PROPSIT con otras variables. Para apoyar la interpretación del puntaje del PROPSIT y su enlace con la teoría de su creación, la relación de las dimensiones del PROPSIT con otras variables es una fuente de evidencias que puede consolidar la creación de estos constructos y su utilidad. El tipo de asociación que generalmente se espera que los constructos y los puntajes directos del PROPSIT cumplan se refiere la convergencia y divergencia, ambos teóricamente guiados y establecidos a priori. En la Figura 5 se resumen los objetivos del presente proyecto.

Objetivos	Tipo de evidencia de validez	Enfoque	Participantes
Objetivo 1a: Identificar la estructura conceptual de los factores psicosociales del trabajo (exposiciones).	Contenido	Cualitativa	Perú
Objetivo 1b: Examinar la claridad, y relevancia del contenido del PROPSIT – FPST	Contenido	Cualitativa – cuantitativa	Perú
Objetivo 1c: Examinar estructura de respuesta de los ítems y validez con criterios externos	Contenido	Cualitativa – cuantitativa	Perú
Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST	Procesos de respuesta	Cuantitativa	Perú
Objetivo 3a: Identificar la estructura dimensional del PROPSIT – FPST.	Estructura interna	Cuantitativa	Perú

Objetivo 3b: Obtener evidencias de validez convergente y divergente interna del PROPSIT.	Estructura interna	Cuantitativa	Perú
Objetivo 3c: Examinar la equivalencia de medición del PROPSIT – FPST.	Estructura interna	Cuantitativa	Perú y México
Objetivo 3d. Examinar la confiabilidad de los puntajes del PROPSIT – FPST	Estructura interna	Cuantitativa	Perú
Objetivo 4. Examinar la relación de los constructos del PROPSIT con otras variables	Relación con otros constructos	Cuantitativa	Perú

Figura 5

Resumen relacional de los objetivos y las evidencias de validez obtenibles

3. Método

3.1. Horizonte teórico-metodológico del investigador

Debido a que la metodología mixta, por definición, involucra técnicas cualitativas y cuantitativas para la recolección y/o interpretación de la información recolectada, se consideró necesario exponer el horizonte teórico-metodológico del investigador, como un contenido que se puede o debería explicitar (Torres-Frías, Moreno-Bayardo, & Jiménez-Mora, 2018; Arias & Giraldo, 2011) en el proceso de investigación. Esto es razonable debido que la implicación o involucramiento personal que el investigador expresada durante la ejecución de la metodología cualitativa es un fenómeno habitual y necesario (Varela & Vives, 2016). Dentro de este involucramiento, se incluyen no únicamente creencias y valores, sino también sus competencias teóricas y metodológicas (Arias & Giraldo, 2011). Ambas áreas de involucramiento pueden ocurrir en diferentes grados (Merino-Soto & Pérez, en preparación). Por lo tanto, se describe el horizonte teórico-metodológico del autor de la presente tesis, que consiste en la acumulación de historias personales de desarrollo teórico y metodológico que el investigador posee, y que influyen el alcance de su comprensión y acciones sobre el objeto de estudio (Merino-Soto & Pérez, en revisión). El investigador del presente proyecto mantiene una línea de

trabajo de más de 15 años en la docencia de pregrado y la investigación científica dentro de las ciencias conductuales, específicamente en psicología. El enfoque metodológico predominante de toda su investigación desde el inicio ha sido el enfoque cuantitativo, aplicado a estudios de validación psicométrica de medidas psicológicas, al desarrollo de propuestas metodológicas en este mismo enfoque, y a estudios sustantivos (no validación psicométrica). Profesionalmente, he tenido eventual práctica clínica en los primeros años de mi ejercicio profesional, pero la mayor experiencia de trabajo ha sido la docencia, y con progresivo incremento en la investigación científica. De acuerdo con lo anterior, las limitaciones de asimilar completamente, y aplicar enfoques cualitativos son claras. Sin embargo, un contrapunto para esta limitación es la orientación por la flexibilidad en la elección de enfoques, así como una visión pragmática (en el sentido de valorar las consecuencias provenientes de acciones con metodologías plurales; Hamui-Sutton, 2013; Johnson & Onwuegbuzie, 2004; Lavelle, Vuk, & Barber, 2013) para orientarse a la solución de problemas. Como consecuencia de esta flexibilidad, se incluye la metodología cualitativa en el desarrollo del presente proyecto, con la cual se lograría no solo profundizar en los resultados de uno de los objetivos propuestos, sino también en orientarse hacia una visión integrativa de diferentes metodologías.

3.2. Diseño de investigación

3.2.1. Presunciones

Para contextualizar metodológica y epistemológicamente, la presente investigación se orienta por las siguientes presunciones. Primero, de acuerdo con la descripción heurística de Kuhn (1962, pag. 66), el problema planteado consiste en la articulación de los hechos

con el modelo teórico. En este sentido, aquí los hechos son aspectos del propio participante (e.g., las conductas o cogniciones auto-informadas) y de sus condiciones laborales (factores psicosociales del trabajo); el modelo teórico es la propuesta de constructos y relaciones entre tales constructos incorporados en el PROPSIT, el cual es el modelo de medición para describir los factores psicosociales del trabajo. Por lo tanto, la presunción es que estos aspectos del participante y de sus condiciones laborales efectivamente existen y pueden ser modeladas por una estructura de constructos.

Segundo, la presunción de que los constructos elaborados en el modelo teórico del PROPSIT pueden ser exitosamente comprendidos desde un enfoque *etic*. Como se describió anteriormente, la interpretación *etic* enfatiza la posible universalidad de los atributos entre las culturas, y las estrategias del diseño de investigación buscan esta representación generalizadora (Berry, 1989, 1989). Esta presunción será probada mediante el diseño *etic derivado*, que recurre a una metodología iniciada en un enfoque *etic*, para verificar lo generalizable de los constructos en el grupo específico, luego se continua con una evaluación *emic* para identificar las diferencias y generar posibles constructos locales, y finalmente se avanza nuevamente hacia un enfoque *etic*, para identificar patrones de contenido que representan interpretaciones similares entre los grupos comparados.

Tercero, otra presunción concierne a las pequeñas diferencias en la metodología de recolección de datos, en que la modalidad principal de recolección (vía web) produciría diferencias psicométricas cercanas a cero frente a la aplicación presencial (e.g., lápiz-papel). El corolario de esto es que la modalidad de aplicación no sería una importante fuente de variación en la respuesta a los datos, debido a la eventual familiaridad de los participantes con los formatos de tipo cuestionarios de auto-reporte. Esta familiaridad puede reducir la carga cognitiva de los participantes. Los estudios que comparan ambas

modalidades han obtenido resultados mixtos, resaltando las ventajas y desventajas de ambas (e.g., Martínez-Gómez, Marín-García, & Giraldo-O'Meara, 2017); sin embargo, hay una tendencia a obtener equivalencia psicométrica entre ellas (Cronck, & West, 2002; Martínez-Gómez et al., 2017).

3.2.2. Diseño de la investigación

Aspectos generales. Para elegir el diseño del presente estudio, se hará un resumen de los objetivos y metas que servirá como argumento y fuente de enlace con la literatura metodológica sobre el diseño más apropiado. Desde una perspectiva general, este estudio es multifásico y secuencial, y se desarrollará en dos etapas, la *fase piloto* y la *fase principal*. En la primera etapa se persigue la meta de explorar las características de contenido del PROPSIT en los trabajadores peruanos, y con ello obtener evidencias de su validez de contenido.

Ya existen evidencias publicadas relevantes para definir la validez de contenido del PROPSIT y de su desarrollo para la población de trabajadores mexicanos (Camacho, 2019; Juárez-García, & Flores-Jiménez, 2020; Juárez-García, Flores, & Imke, 2019), mientras que en trabajadores peruanos este instrumento es completamente nuevo, sin ninguna investigación empírica. Por lo tanto, el presente proyecto se enfocó en la adaptación del PROPSIT a los trabajadores peruanos.

En el presente proyecto, la validez de contenido será evaluado respecto a los criterios de relevancia, claridad y validez con criterios externos, aspectos que usualmente son conocidos como argumentos provenientes del análisis lógico y empírico (American Psychological Association et al., 1999). En alguno de estos análisis, se obtiene el juicio del evaluador o participante sobre el criterio de contenido (para el presente estudio, relevancia,

claridad), y, por lo tanto, el juicio evaluativo es altamente requerido. En el presente estudio se planificó obtener este juicio evaluativo por medio de la metodología cuantitativa. La estrategia de recolección de datos utilizará preguntas abiertas y cerradas, agrupados en un cuestionario a priori estructurado (Creswell, 2013), de una manera que facilite la obtención directa de la información cuantitativa o la transformación de información textual (originada en las preguntas abiertas) hacia la información cuantitativa. El análisis e interpretación final de esta información será cuantitativa. Por lo tanto, el diseño de la etapa piloto puede ser identificado como un diseño de investigación mixta.

Por otro lado, en la segunda fase del estudio (etapa principal), el énfasis es cuantitativo en la recolección, análisis, y elaboración de inferencias. En esta etapa se buscan dos cosas: evaluar estadísticamente las propiedades psicométricas del PROPSIT, y el grado de similaridad de las propiedades psicométricas del PROPSIT en trabajadores peruanos y mexicanos.

Identificación del diseño. Desde hace algunas décadas, se han elaborado numerosas taxonomías de los diseños de investigación mixta, que varían en complejidad y en la inclusión de criterios importantes para definirlos (Bentahar, & Cameron, 2015; Leech, & Onwuegbuzie, 2007). Una taxonomía relevante para el presente estudio es una que establece tres criterios (Bentahar, & Cameron, 2015; Leech, & Onwuegbuzie, 2007), relacionados con a) el orden de presentación de la metodología cuantitativa y cualitativa (criterio de tiempo), b) la saliencia del método cuantitativo o cualitativo (criterio de dominancia o estatus), y c) el grado de combinación de ambos métodos (criterio de magnitud de la combinación). El primer criterio (tiempo) indica la aplicación concurrente o secuencial de los métodos, el segundo criterio (saliencia) apunta al método dominante aplicado en la recolección de datos y la presentación de resultados, de tal modo que los

métodos (cuantitativo o cualitativo) puede poseer igual o diferenciado (dominancia) estatus. Finalmente, el criterio de magnitud de la combinación se refiere al grado en que ambos métodos están diseminados en los diferentes componentes de la investigación (respectivamente, completamente mixto o parcialmente mixto), es decir, en los objetivos, recolección de datos, análisis, e interpretación, o centrados en una parte del mismo.

De acuerdo a lo anterior, el diseño general del presente proyecto corresponde con el diseño de investigación *parcialmente mixta* (Leech, & Onwuegbuzie, 2009), porque la combinación de metodologías no se implementará en todas las fases del estudio, sino en la primera fase (correspondiente a la validez de contenido), mientras para la segunda fase, se implementará la metodología cuantitativa. Con respecto al diseño específico en cada fase, en la primera fase se implementará un *diseño mixto concurrente*, en que la recolección de datos con ambas metodologías se inicia simultáneamente, pero dado que el método predominante es cuantitativo, éste puede ser denominado *concurrente anidado* (Creswell, 2013); adicionalmente, la metodología predominantemente de esta fase es cuantitativa (Johnson, Onwuegbuzie, & Turner, 2007). La implementación de este tipo de diseño mixto en esta fase persigue un objetivo *expansivo*, en que la aplicación de los diferentes métodos se asocia a las diferentes metas específicas (Johnson, & Onwuegbuzie, 2004); por ejemplo, en el presente estudio, la metodología cuantitativa se usará para obtener el juicio de la relevancia, claridad e validez con criterios externos del contenido del PROPSIT, mientras que la información cualitativa se usará para obtener información textual de los factores estresores y factores positivos en el trabajo. Como ambos están interrelacionados dentro de la misma fase, son adecuados para un enfoque mixto (Johnson et al., 2007).

Para la segunda parte del estudio (fase principal), desde una perspectiva de metodología única, específicamente cuantitativa, se usó el sistema de clasificación de investigaciones empíricas propuesto por Ato, López y Benavente (2013). Por lo tanto, el diseño en esta fase corresponde esencialmente a una investigación instrumental mediante la metodología cuantitativa.

3.3. Participantes

Elegibilidad. En la etapa piloto del presente estudio, los participantes elegibles deben cumplir con los siguientes criterios de inclusión aplicables a ambos países: a) ser legalmente adulto, b) estar activamente trabajando como dependiente en una empresa u organización formal, c) tener un tiempo mínimo de trabajo activo de seis meses, c) ser de nacionalidad peruana o mexicana de nacimiento, d) haber logrado escolaridad secundaria mínima, y e) aceptar la participación voluntaria y anónima en el estudio. El criterio de exclusión será: a) el propio retiro del participante durante la recolección de la información, y b) puntajes bajos en el índice de dominio del idioma. Éste último criterio específico ayudará a controlar el efecto del nivel del dominio del idioma sobre los juicios de claridad de los ítems.

En la etapa principal del estudio, se aplicarán los mismos criterios de inclusión. Los criterios de exclusión se aplicaron de acuerdo con lo siguiente: la cantidad de información demográfica no respondida (más de la mitad de estos ítems), la cantidad de ítems no respondidos en los cuestionarios (dependiendo de la extensión del instrumento, pero en general, 3 o más ítems sin responder), la prevalencia individual de patrones de respuesta irrelevantes (determinado por criterios cuantitativos de los métodos utilizados), y el grado

de instrucción potencialmente limitante para responder a los cuestionarios (por ejemplo, nivel de primaria elemental o nivel de secundaria incompleta). Un elevado número de registros en alguno de los criterios de exclusión indicará al participante como candidato para removerlo, debido a que hacer una imputación de esta información ausente reducirá la varianza de las variables imputadas, y conduce a crear patrones grandes de respuestas ficticias.

En esta misma etapa principal, los participantes serán también distinguidos respecto a su área laboral formal. El tipo de profesión inicialmente proyectado como participantes del presente estudio serían trabajadores en un área relativamente homogénea de trabajo. Por ejemplo, se proyecta que los participantes sean a) trabajadores empleados de empresas formales, y b) docentes de educación básica. Ambos son elegibles porque el acceso a ellos es más viable en relación con los recursos del investigador, y son más próximos en la red de contacto del investigador. Adicionalmente, existen experiencias previas de contacto exitoso en la investigación (Fernández et al., 2015; Fernández-Arata & Merino-Soto, 2016) que fortalecen esta decisión sobre la muestra.

Tamaño muestral. En la etapa piloto se incluirán trabajadores o participantes-objetivo, y se aceptará un mínimo aproximado de 25 participantes, con un rango recomendado de entre 30 y 40 participantes (Hertzog, 2008). Estos participantes tendrán características heterogéneas respecto al trabajo realizado, para poder evaluar la variabilidad de los resultados en la validez de contenido, y potencialmente facilitar la transferencia de los resultados entre las áreas laborales participantes. En esta misma etapa piloto, también se elegirán jueces expertos, y se estima un tamaño alrededor de 10 jueces como mínimo (Angulo-Ramos, Córdova-Delgado, & Merino-Soto, 2021). Los jueces expertos serán identificados y convocados a participar de acuerdo con la disponibilidad, y respecto a los

siguientes criterios (Urrutia, Barrios, Gutiérrez, & Mayorga, 2014): experiencia profesional en psicología organizacional, docencia en temas relevantes, investigación (publicada o no), y grados académicos alcanzados.

Para estimar el tamaño muestral del estudio principal, se tomó en cuenta la metodología principal de análisis, el cual será el modelamiento de ecuaciones estructurales (SEM). Siguiendo los parámetros de Westland (2010), se definieron varios criterios previos para la estimación del tamaño muestral mínimo, como el poder estadístico y nivel de probabilidad mínimos deseados (convencionalmente, 80% y $p < .05$, respectivamente; Cohen, 2008), la magnitud del parámetro principal (en este estudio, la carga factorial, $\lambda = .40$), el número de variables latentes (se consideró dos variables latentes mínimos en los predictores, mediadores y criterios; en total, 6), y el número de variables observadas (10 ítems en las dimensiones de alto orden). Usando un programa ad hoc (Soper, 2017), el mínimo tamaño muestral con estos parámetros a priori fue 489; sin embargo, como se efectuarían estudios de invarianza, este se duplicó para designar el tamaño muestral en cada muestra comparada. Por lo tanto, el tamaño mínimo total aproximado estaría alrededor de 978 participantes.

3.4. Instrumentos

Encuesta de Procesos Psicosociales del Trabajo (PROPSIT; Juárez, 2017; Juárez & Camacho, 2011). Es el instrumento que operacionalizará el modelo de evaluación de procesos psicosociales en el trabajo para el presente estudio. Se presenta en la modalidad de cuestionario auto-administrado, el cual es un método ampliamente elegido en la investigación aplicando la metodología de encuestas (Holbrook, 2008) por cuestionarios, y en particular en la investigación psicológica (Vallejo-Medina, et al., 2017).

Desarrollo. La construcción del PROPSIT fue creada para la población de trabajadores mexicanos, en base al análisis del estado del arte sobre las teorías existentes de los factores psicosociales del trabajo (específicamente, el modelo de demandas-recursos), pero vinculada con la evidencia acumulada interculturalmente sobre los constructos que recurrentemente participan en el contexto laboral y en la respuesta del trabajador. Fue creada como herramienta para la investigación científica y para el diseño de intervenciones psicosociales ocupacionales. En la sección del Marco Sustantivo, se describe las bases teóricas de su construcción.

Estructura de respuesta. Con el propósito de maximizar la variabilidad de respuesta y detectar finas graduaciones en la experiencia de los examinados, la escala de respuesta para los ítems del PROPSIT es ordinal de siete puntos, desde 0 (*Nunca*) hasta 6 (*Siempre*), distribuidos simétricamente alrededor del punto medio, 3 (*Regularmente*). Este rango de opciones es similar a otras medidas involucradas en la salud ocupacional, por ejemplo, el MBI (Maslach et al., 1996), el UWES (Schaufeli et al., 2002), medidas de prolongada fatiga (Bültmann et al., 2000), y desafío-obstáculo (LePine, LePine & Jackson, 2004), entre otros. La extensión del escalamiento de respuesta está en la amplitud recomendada por estudios psicométricos (Finstad, 2010), y se encuentra en el rango de las opciones comúnmente utilizadas en la investigación psicológica (entre cinco y siete opciones de respuesta). El rango de experiencia muestreada en las opciones de respuesta es de frecuencia, etiquetada de dos maneras para asegurar la comprensión del respondiente: una etiqueta única, y su correspondiente definición ultra-breve. Por ejemplo, para la opción 1, aparece, *Rara vez (muy pocas veces, al año o menos)*”. Esta definición adiciona claridad a la etiqueta y también mejor diferenciación entre las opciones de respuesta.

Presentación. Originalmente, el PROPSIT se expone al examinado luego del llenado del consentimiento informado y de la hoja demográfica. Esta presentación de los ítems empieza con una declaración del enfoque que se espera del examinado para responder (*No hay respuestas buenas ni malas. Es importante que conteste absolutamente todas las frases de una manera sincera, no deje alguna sin contestar*); luego aparece el escalamiento de respuesta ordinal de siete puntos, y continua con una explicación del proceso de respuesta (*Utilizando la escala de anterior, marque el número que corresponda para indicar con qué frecuencia ha experimentado estos sentimientos o situaciones. Si nunca lo ha experimentado, marque 0. Si en efecto ha experimentado este pensamiento o sentimiento, marque la respuesta que mejor lo describa.*). Finalmente, se continúa con los ítems.

Estructura conceptual. El PROPSIT describe tres niveles de interpretación de alto orden de interpretación: *Factores Psicosociales del Trabajo* (FPT), *Efectos Psicosociales del Trabajo* (EPT), *Proceso Salud-Enfermedad* (PSE), y uno complementario, *Factores Extra-laborales* (FET). En otro orden de interpretación, cada uno de estos niveles subsume a un segundo nivel de interpretación, correspondiente con la *valencia* de los niveles anteriores. En FPST se evalúan los *Riesgos Psicosociales* (valencia negativa) y los *Recursos Favorables* (valencia positiva); en EPT se evalúan los efectos Positivos y Negativos; y en PSS se evalúan la *Percepción de Enfermedad* (valencia negativa) y la *Percepción de Salud* (valencia positiva). Los FET exploran la *Falta de Control Psicológico*, la *Sobre-implicación al trabajo*, *Exigencias laborales* y *Conflicto Trabajo-Familia*. En otro nivel de interpretación, cada uno de estos aspectos muestrea indicadores conceptuales con mayor nivel de especificidad. Por ejemplo, la dimensión *Recompensas* y

Desarrollo Laboral muestrea indicadores conceptuales de: *justicia organizacional, salario motivante, reconocimiento por el trabajo, posibilidades de desarrollo laboral y profesional (estatus), seguridad en la conservación del empleo, y tarea gratificante y con trascendencia*. En la Figura 6 se muestran los constructos, así como como sus niveles de interpretación e intervención asociados. Para el presente estudio, sin embargo, se utilizó la sección de los factores psicosociales del trabajo, debido que el foco del presente proyecto son estos factores, mientras que el resto del PROPSIT evalúa otros aspectos vinculados, pero conceptualmente diferentes.

Hoja demográfica. Los participantes fueron caracterizados de acuerdo con la información general demográfica y a las características laborales de trabajo. En el primero se registró el sexo, la edad cronológica, personas con quien convive, número de hijos, (que viven en el hogar y con dependencia económica), personas mayores dependientes económicamente que viven en el hogar estado civil, tipo de contribución financiera al hogar, nivel educativo y actividad de estudio actual. Respecto a la caracterización laboral, esta será obtenida con ítems que recojan información sobre el tipo de contrato, años de experiencia laboral en la empresa actual, años de experiencia laboral anterior y antes de ingresar a la empresa actual, puesto de trabajo, clasificación del puesto de trabajo, número de horas de trabajo semanales, empleo reenumerado adicional, horas semanales a otros empleos reenumerados. En el Apéndice sobre los instrumentos, se presenta la estructura y especificaciones de la hoja demográfica. Este formulario de preguntas se aplicará en la etapa principal del estudio.

Factores Psicosociales del Trabajo (enfoque de prevención o promoción primaria)	Efectos psicosociales del trabajo (prevención secundaria)	Proceso Salud-Enfermedad (prevención terciaria)	Otros Factores Psicosociales extra-laborales
<p>Recursos favorables (valencia positiva)</p> <p>Recompensas y desarrollo laboral</p> <p>Justicia organizacional Salario motivante</p> <p>Reconocimiento por el trabajo</p> <p>Possibilidades de desarrollo laboral y profesional (estatus) Seguridad en la conservación del empleo</p> <p>Tarea gratificante y con trascendencia</p> <p>Control laboral y contenido de la tarea</p> <p>Influencia-autonomía en el trabajo</p> <p>Utilización de habilidades en el trabajo</p> <p>Trabajo variado (no monótono)</p> <p>Funciones y roles claros</p> <p>Recursos para realizar el trabajo</p> <p>Recursos materiales, equipos, herramientas para el trabajo Capacitación, adiestramiento</p> <p>Clima y apoyo social en el trabajo</p> <p>Apoyo de compañeros Apoyo de jefes</p> <p>Clima de unión</p> <p>Congruencia de valores</p> <p>Congruencia de valores individuo-organización</p> <p>Congruencia de valores individuo-compañeros-superiores</p>	<p>Bienestar y Entusiasmo (valencia positiva)</p> <p>Energía</p> <p>Satisfacción con el trabajo</p> <p>Entusiasmo</p> <p>Comodidad</p> <p>Absorción</p>	<p>Percepción de salud (valencia positiva)</p> <p>Salud física positiva percibida</p> <p>Sueño reparador Sentimientos de competencia</p> <p>Felicidad y afectos positivos</p>	<p>Falta de control psicológico</p> <p>Control emocional inadecuado</p> <p>Creencias de control inadecuadas</p> <p>Control en las relaciones sociales</p> <p>Sobre implicación/adicción al trabajo</p> <p>Conflicto familia/trabajo</p> <p>Por conflictos emocionales con la pareja o familia Por exigencias de tiempo o de actividades demandantes Por indiferencia familiar o de la pareja a la ética del trabajo Por exigencias financieras de la familia</p> <p>Otras exigencias extra-laborales Otros estresores extra-organizacionales</p>
<p>Clima y apoyo social en el trabajo</p> <p>Apoyo de compañeros Apoyo de jefes</p> <p>Clima de unión</p> <p>Congruencia de valores</p> <p>Congruencia de valores individuo-organización</p> <p>Congruencia de valores individuo-compañeros-superiores</p>	<p>Malestar y Burnout (valencia negativa)</p> <p>Agotamiento-fatiga</p> <p>Hastío psíquico (desagrado-indiferencia)</p> <p>Tristeza</p> <p>Nerviosismo</p> <p>Presentismo: Desvinculación del desempeño efectivo por enfermedad o malestar.</p>	<p>Percepción de enfermedad (valencia negativa)</p> <p>Síntomas somáticos:</p> <p>Trastornos mentales y de adaptación:</p> <p>Otros trastornos mentales:</p>	
<p>Riesgos Psicosociales (valencia negativa)</p> <p>Demandas laborales</p> <p>Carga y ritmo de trabajo Alta responsabilidad y peligrosidad Jornadas, turnos u horarios Cognitivas o atencionales Emocionales (trato con personas) Esfuerzo físico Ambiente físico</p> <p>Acoso laboral (psicológico)</p> <p>Por superiores Por compañeros o subordinados</p> <p>Liderazgo estresante</p> <p>Supervisión controlante</p> <p>Retroalimentación negativa/ inadecuada del supervisor</p>			

Figura 6

Estructura conceptual y operacional del PROPSIT

El PROPSIT es un extenso instrumento que puede ser utilizado en su totalidad o parte de ella, para los propósitos definidos por el usuario. Debido que la investigación del presente se enfocó que las exposiciones laborales, el área de contenido utilizado para la validación en la muestra peruana fue la de Factores Psicosociales del Trabajo.

Escala de sobrecarga de estrés – breve (SOS-S; Amirkhan, 2016). Medida breve que evalúa la percepción de sobrecarga de estrés con 10 ítems, expresado en dos dimensiones, vulnerabilidad personal y carga de eventos (5 ítems en cada uno). Resultados de validez en otros estudios (Duan, & Mu, 2018; Wilson et al., 2018) indican que el puntaje total puede ser estimado para representar la percepción de sobrecarga de estrés.

Escala de autoeficacia ocupacional (OSES; Rigotti, et al., 2008). Instrumento de seis ítems que evalúa la autoeficacia en el contexto laboral global, mediante la percepción de suficiencia y confianza en la habilidad para eventos laborales. En recientes validaciones peruanas, se corroboró la alta magnitud de confiabilidad y relación con criterios laborales y otros constructos (Merino-Soto, Lima-Mendoza, Lozano-Huamán, Calderón de la Cruz, & Juárez-García, 2021).

Escala de Engagement de Utrech (UWES-3; Schaufeli, Shimazu, Hakanen, Salanova, & De Witte, 2019). Evalúa el entusiasmo laboral, con tres ítems sobre vigor, dedicación y absorción elegidos con base racional y empírica por los autores y derivados del UWES-9. El escalamiento de respuesta es de siete puntos, entre 0 (*nunca*) y 6 (*siempre*). En dos estudios peruanos, los resultados de validez confirmaron su validez de constructo respecto a la estructura interna, confiabilidad y relación con otras variables (Calderon-De la Cruz, Merino-Soto, Reyes-Ayala, & Luna-Ayón, 2021; Merino-Soto et al., Toledano-Toledano, 2022).

Formulario para la Validez de Contenido (FVC; Merino-Soto, 2018). Este formulario de preguntas se aplicará en la etapa principal del estudio. Su descripción aparece en el Objetivo 1b. Examinar la claridad, relevancia y validez externa del contenido del PROPSIT – FPST.

Medidas de ítem único. Existe numerosa evidencia científica que señala a las mediciones por medio de ítems únicos como razonables indicadores de sus constructos psicosociales (Bergkvist, 2014; Bergkvist & Rossiter, 2007; Diamantopoulos, Sarstedt, Fuchs, Wilczynski & Kaiser, 2012; Fuchs & Diamantopoulos, 2009; Wanous, Reichers, & Hudy, 1997), pero condicionales al propósito de su uso; por ejemplo, la recolección de datos en el marco de una investigación epidemiológica, o evaluaciones de tamizaje (screening) en poblaciones que requieren brevedad de tiempo para la aplicación de encuestas. Esto sugiere que medidas de ítems únicos no son intrínsecamente precarios, aun cuando sus limitaciones psicométricas se han discutido en la literatura. Por lo tanto, este proyecto incluirá este tipo de medidas para corroborar la validez del modelo evaluativo propuesto.

Ítem Único de Estrés (Elo, Leppänen & Jahkola, 2003). Está destinado para evaluar la intensidad del estrés en un contexto general o de la vida diaria. El respaldo para su validez tiende a ser favorable (Elo et al., 2003; Littman, White, Satia, Bowen & Kristal, 2006; Salminen, Kouvonen, Koskinen, Joensuu & Väänänen, 2014). En población peruana, existen resultados psicométricos satisfactorios aún en desarrollo (Merino-Soto, 2017). Este instrumento será covalidado en el presente estudio junto al modelo evaluativo propuesto. El contenido del ítem se presenta en el Apéndice.

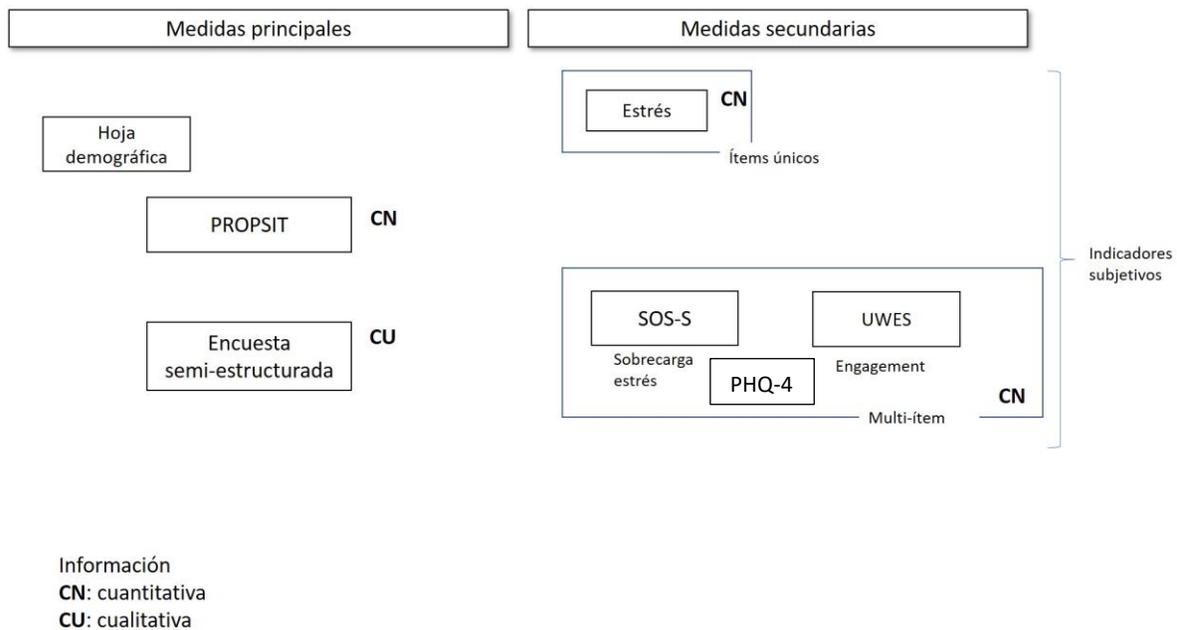


Figura 7
 Organización de los instrumentos del estudio

Encuesta semi-estructurada. La aplicación de procedimientos cualitativos para la obtención de información será en dos momentos: primero, durante la evaluación de la validez de contenido, y segundo, durante la profundización de los resultados obtenidos por la metodología cuantitativa. En la primera aplicación, correspondiente con el marco de la validez de contenido, se exploró mediante una *encuesta semiestructurada* los estresores laborales que caracterizan la vida del trabajador. Se eligió este método porque se focalizará directamente sobre una estructura de contenido correspondiente con la propuesta evaluativa del PROPSIT; esta estructura puede identificarse como las hipótesis o la estructura pre-definida de conceptos contra los que se comparan los conceptos emergidos de la entrevista.

Esta encuesta semiestructurada contiene dos partes; la primera comprende dos preguntas abiertas y no dirigidas, en que se indaga directamente sobre los estresores laborales y las condiciones generadoras de bienestar. Para iniciar la encuesta, el participante

primero llenó datos sociodemográficos, y luego, en el contenido objetivo de la encuesta, las instrucciones incluidas le indujeron a que escriba en una hoja las situaciones que le generan estrés en el trabajo. Como parte de las instrucciones, se presentó una definición de estrés laboral, y que, para cada situación estresante declarada, califique la frecuencia con que ese estresor aparece en su vida laboral. En la siguiente parte de esta encuesta, se les solicitó que identifiquen los aspectos de su trabajo que más le gustan; y como en la indagación anterior sobre las situaciones estresantes, también se solicitó que califique la frecuencia con que ocurren estos aspectos positivos. La escala de calificación se planteó sobre un rango de tres opciones: *pocas veces*, *frecuentemente* y *muchas veces*. Este escalamiento breve es parsimonioso, diferencia en categorías grandes la frecuencia de las respuestas obtenidas, y puede ser mejor aceptada por los entrevistados (Flores & Juárez-García, 2016). Esta metodología fue aplicada en general en un estudio previo, y corresponde a una aproximación con la metodología mixta de recolección y análisis de información (Flores & Juárez-García, 2016). La metodología aplicada en esta parte del estudio está influenciada por el método propuesto por Juárez-García (Flores & Juárez-García, 2016; Juárez-García, Flores-Jiménez, & Pelcastre-Villafuerte, 2020; Juárez-García, Camacho-Ávila, García-Rivas, & Gutiérrez-Ramos, 2021), donde los factores de valencia negativa (de riesgo psicosocial) y positiva (recursos favorables) con explorados con componentes cualitativos y cuantitativos.

3.5. Procedimiento

El presente estudio es multifásico, porque está estructurado en varias fases organizadas e interdependientes, y dentro de cada fase, existen grupos de procedimientos conducidos simultánea o secuencialmente en relación con el objetivo específico. Las dos

grandes fases del estudio son la fase *piloto* y la fase *principal*; ambas están vinculadas en la manera que los resultados de los estudios realizados en la fase piloto influenciarán directamente en las decisiones estructurales de la fase principal. Esta influencia se debe que la fase piloto abordará la validez de contenido, y por lo tanto la conservación o modificación del contenido del PROPSIT será decidida en esta fase. Ambas fases encadenadas poseen algunos aspectos metodológicos diferentes.

3.5.1 Enfoque general

Muestreo. En general, el presente estudio usará un enfoque de muestreo no probabilístico, en que la selección de las unidades no requiere algún procedimiento aleatorizado que permita estimar la probabilidad de los parámetros obtenidos, así como la probabilidad de selección del participante (Battaglia, 2008a). Este tipo general de muestreo tiene la ventaja de controlar aspectos del proceso de elección de los examinados y sus contextos de pertenencia, enfatizando la coherencia del proceso investigado en lugar de la generalización de los resultados. También, tiende a ser un procedimiento eficiente en costo, recursos y tiempo (Battaglia, 2008a), y especialmente útil cuando se planifica un tamaño muestral reducido (Black, 2010). Más que la excepción, los métodos de muestreo no probabilístico son una tendencia consistente en la investigación en muchas áreas disciplinares de las ciencias (Elliott, & Valliant, 2017); y aplicando moderaciones a la interpretación de los resultados obtenidos en ellas, la sustantividad o importancia de los resultados aún pueden mantenerse.

Recolección de información. En general, las respuestas de los participantes serán obtenidas mediante la aplicación *presencial directa examinado-examinador*, y

la *aplicación mediada por web*. Ambas modalidades serán utilizadas en la etapa piloto o en la etapa principal. Respecto a la aplicación mediada por web, hasta la fecha de hoy, hay suficiente información empírica sobre la aplicación mediada por web, específicamente en la investigación que usa cuestionarios, y en casi todas las disciplinas científicas, incluidas las relevantes al presente estudio, esto es, psicología, medicina, salud pública y epidemiología. Aunque existen hallazgos contradictorios reconocidas en la literatura, la tendencia general empírica fuertemente señala que la aplicación mediada por web produce resultados confiables, válidos, aceptablemente representativos, eficientes en costo económico y tiempo, con mayor cobertura de los participantes encuestados, mejor protección de la anonimidad de respuesta, menor tasa de no-respuesta, y aceptable equivalencia frente a la aplicación tradicional (Campos, Zucoloto, Bonafé, Jordani, & Maroco, 2011; Denscombe, 2009; Liao, & Hsieh, 2017; Meyerson, & Tryon, 2003; Sax, Gilmartin, & Bryant, 2003; van Gelder, Bretveld, & Roeleveld, 2010). Estos resultados se mantienen en temáticas y disciplinas (por ejemplo, epidemiología y salud pública) relevantes al contexto temático de los factores psicosociales del trabajo (Campos et al., 2011; Hohwü et al., 2013; van Gelder et al., 2010). No habrá tiempo límite establecido para el llenado de los instrumentos, y se mantendrán instrucciones similares de respuesta en la introducción de la encuesta, y en cada instrumento. El orden de presentación de los instrumentos será el mismo en ambas fases.

Por otro lado, la aplicación presencial directa corresponde al procedimiento tradicional de entrega de instrumentos; esta se hará mediante un examinador que explicará las instrucciones de un modo estandarizado e invariable en los diferentes

grupos muestreados. La aplicación presencial expondrá las instrucciones enfatizando la honestidad de respuesta, la cumplimentación de todas las cuestiones, el cambio de respuesta según sea necesario, la lectura cuidadosa de las instrucciones, la condición de anonimato, la prestación de ayuda para resolver dudas, y la confirmación de la participación voluntaria mediante la firma del formulario de consentimiento informado. La elección de la aplicación online o presencial en la etapa principal del estudio estará guiada por la oportunidad de utilizarlas en el contexto específico.

Evaluación de propiedades psicométricas. Como el uso de los resultados tendrán dos líneas de aplicación, una teórica y otra práctica, ambos requieren información precisa y asociada muy cercanamente a los constructos intentados para medir. Por lo tanto, todas las medidas involucradas serán evaluadas psicométricamente, y no únicamente el PROPSIT.

En el presente estudio, los análisis correspondientes con la evaluación de las propiedades psicométricas se organizaron también en las dos fases principales del estudio; esto es, en la etapa piloto se examinaron las evidencias de validez de contenido, y en la etapa principal, se examinarán las evidencias de estructura interna, la confiabilidad, la invarianza de medición, y asociación con otras variables.

3.5.2 Procedimientos en la etapa piloto

Esta primera fase actuó como pre-prueba, o las pruebas iniciales del PROPSIT en los participantes peruanos. El objetivo en esta fase fue la implementación de los procedimientos para obtener evidencias de la validez de contenido del PROPSIT, tal como

fueron descritos en la sección Evaluación de propiedades psicométricas; esto se refiere a probar la relevancia, la claridad o comprensión.

Muestreo de los estudios pilotos. El modo específico de muestreo en la fase piloto será de tipo *intencional*, en que la elección de los participantes se sustenta en el conocimiento juicioso de la población por parte del investigador, para explorar las variables de interés (Battaglia, 2008b; Etikan, Musa, & Alkassim, 2016; Patton, 1990). Esto requiere un componente subjetivo para elegir a los participantes, por lo que la validez de este descansa sobre una base lógica más que probabilística (Lavrakas, 2008). El tamaño muestral generalmente es pequeño para los procedimientos aplicados en el piloto, sea para la investigación científica o clínica (Etikan et al., 2016), y descansa sobre la evaluación inicial de las características del instrumento elegidas por el investigador. Aproximadamente, 25 y entre 30 y 40, respectivamente (Hertzog, 2008) para los participantes, y alrededor de 10 expertos (Angulo-Ramos, Merino-Soto, & Ramón-Rojas, 2021).

Recolección de información de los estudios piloto. Fue aplicada presencialmente o mediada por web, de acuerdo con la situación que maximice el número óptimo de participantes. Se usó el Formulario de para Validez de Contenido (ver Apéndice). En cada grupo aplicado, se mantuvieron invariantes las instrucciones de aplicación.

Análisis de la información recolectada en la etapa piloto. Esencialmente, se efectuaron análisis para obtener evidencias de validez de contenido. El contenido del instrumento fue evaluado con énfasis en la representación del contenido (la representación de los ítems en relación a los constructos intentados para representar), la claridad de los ítems (grado de claridad o entendibilidad), y relación

de los ítems con variables externas (asociación de los ítems con constructos y criterios externos al modelo de medición). El presente estudio eligió a participantes y jueces expertos para evaluar estos aspectos, y sus respuestas fueron cuantificados para cada ítem y para el conjunto unitario de ítems (Polit & Beck, 2006; Polit, Beck & Owen, 2007) mediante el coeficiente V (Aiken, 1980, 1985). Su significancia estadística evaluada mediante intervalos de confianza asimétricos (Merino & Livia, 2009; Penfield & Giacobbi, 2004)

Representación conceptual del ítem (relevancia). Usualmente, este aspecto se refiere a la relevancia del contenido del ítem para formar parte de la identificación conceptual del constructo y su expresión en conductas coherentes (Merino & Livia, 2009). Debido que la creación y selección del contenido del PROPSIT fue conducido teóricamente, y su uso fue destinado originalmente al contexto mexicano, para generalizarlo a otros contextos, se requiere evaluar su contenido empíricamente, y aplicar potenciales modificaciones desde una perspectiva de descentración (como fue explicada en la sección introductoria). Para evaluarlo, se solicitó a los jueces expertos que valoren su grado de relevancia respecto al constructo presentado. Esta instrucción enfatizó la evaluación de esta característica para sea independiente de la claridad del ítem. La selección de los jueces fue evaluada con varios criterios temáticamente orientados: investigaciones publicadas, docencia universitaria, práctica profesional (evaluación o intervención), y grado mínimo de maestría. Sus respuestas fueron escaladas en un formato ordinal de cinco opciones, desde “nada representativo” hasta “completamente representativo”.

Claridad de los ítems. Este criterio fue examinado para ver el grado en que el ítem es claro, comprensible o inteligible. Debido a las diferencias que se han detectado entre grupos de calificadores expertos y participantes en la evaluación de claridad de los ítems (Angulo-Ramos, Merino-Soto, & Ramón-Rojas, en revisión; Merino, 2016), se solicitó a los propios participantes para identificar ítems poco claros. Se usó una escala de respuesta ordinal de cinco opciones, desde “nada claro” hasta “completamente claro o entendible”.

Validez de los ítems. Los ítems fueron examinados en su validez con criterios externos, para verificar la capacidad asociativa de cada ítem individual sobre constructos teóricamente implicados.

Análisis de la encuesta semiestructurada

El material proveniente de la encuesta semiestructurada fue registrado textualmente, debido que en la sección donde se presentan las preguntas abiertas, los participantes escribieron sus respuestas. Una vez obtenido el registro textual, esta información fue codificada mediante la identificación de patrones temáticos, y ejecutado por el investigador del proyecto. Esta codificación se hizo mediante un procedimiento selectivo (codificación selectiva; Kvale, 2011), y dentro de un enfoque hipotético-deductivo en el proceso de codificación (Wengraf, 2001), debido que una estructura conceptual ya existe y proviene de la organización conceptual del PROPSIT. Para estimar la consistencia de la codificación, se requirió a un psicólogo con extensa experiencia profesional y otro psicólogo investigador, ambos del área organizacional, que independientemente codifiquen el material. Ambos recibieron información sobre las categorías conceptuales establecidas (i.e., de la organización conceptual del PROPSIT), y sobre las categorías que emergieron en la

primera codificación construida por el investigador del proyecto. La consistencia fue estimada cuantitativamente por coeficientes de acuerdo, corregidos por acuerdo aleatorio.

3.5.3 Procedimientos en la etapa principal

Muestreo del estudio principal. Como se explicó en el enfoque general de la metodología, el muestreo fue no probabilístico. En contraste con la fase piloto, en que el muestreo es intencional, en esta etapa se planifica que el muestreo específico será de conveniencia.

Recolección de información del estudio principal. Para cumplimentar el objetivo de la evaluación de la estructura interna y la invarianza de medición, se aplicó la sección de factores psicosociales del PROPSIT, junto con la hoja demográfica. Esta aplicación fue mediada por una plataforma web, y estuvo condicionado por disponibilidad de recursos de la empresa participante. Ambas modalidades de aplicación mantuvieron las mismas instrucciones y orden de los cuestionarios. Por otro lado, para cumplimentar el objetivo de las relaciones nomológicas del PROPSIT con otras medidas, se tomó en cuenta la extensión del PROPSIT y del conjunto de instrumentos acompañantes. La extensión total de medidas puede ocasionar en los participantes fatiga y actitud evitativa hacia la cumplimentación del paquete de instrumentos. Por lo tanto, se planearon algunas estrategias para maximizar la tasa de respuesta en cada examinado, la misma que interacciono con la situación concreta que encuentre el investigador. La primera estrategia consiste en que un grupo de participantes resuelva todos los materiales elegidos cuando la situación permita a) descanso, b) suficiente tiempo de respuesta,

y c) fuerte apoyo institucional. La segunda, correspondiente con la aplicación online, fue la configuración del formulario para que únicamente permita respuestas a todas las afirmaciones de los cuestionarios.

Análisis de la información recolectada en la etapa principal

Análisis preliminar: calidad de los datos. En primer lugar, se evaluó la calidad de los datos respecto a la *aceptabilidad*, y *patrones de sesgo de respuesta*. En la investigación hispana, especialmente publicada en Latinoamérica, se conoce poco de estudios empíricos o teóricos que expongan hallazgos al respecto, y por lo tanto no se dispone de una línea-base de investigaciones para propósitos comparativos. Por lo tanto, los aspectos mencionados fueron evaluados en antes de los análisis principales del estudio.

Aceptabilidad de respuesta. Fue examinado mediante la tasa de respuesta y valores perdidos (Menard et al., 2014; Netherlands Cancer Institute, Amsterdam, 2002). Estos aspectos se hicieron mediante el porcentaje de no-respuesta en dos niveles: los ítems y los participantes. En el nivel de los ítems, el porcentaje de aceptabilidad fue evaluado respecto a las recomendaciones o estándares empíricos reportados en la literatura. En el nivel de los participantes, se registró el número de ítems no respondidos en cada instrumento multi-ítem, el cual será evaluado contra el número de ítems que contiene el instrumento. En conjunto, la aceptabilidad de respuesta será analizada en las aplicaciones lápiz-papel de los instrumentos, debido que la aplicación online controlará automáticamente el llenado completo. En la situación de un número reducido de ítems sin responder, se aplicó procedimientos de imputación de valores perdidos, y apropiados para datos categóricos y basados en

estimaciones psicométricas (Sijtsma & Van der Ark, 2003; Van Ginkel, & Van der Ark, 2005; Van Ginkel, Van der Ark & Sijtsma, 2007)

Patrones de respuesta. Este aspecto se relaciona con el estilo de respuesta ante los instrumentos, y que es irrelevante al contenido del constructo evaluado. Existe evidencia de este tipo de conducta en las medidas de autoinforme, y que varía interculturalmente (Fisher et al., 2009; Harzing, 2006; Johnson, Kulesa, Cho & Shavitt, 2005; Van Dijk et al., 2009). Parece que la literatura empírica en la evaluación del riesgo psicosocial ha generado poco conocimiento al respecto, pero dado que los instrumentos que se utilizarán en este estudio son de auto-reporte, se puede plantear que estos problemas existirán en algún grado.

Estructura interna

Se aplicará un análisis factorial confirmatorio desde el modelamiento de ecuaciones estructurales (*Structural Equation Modeling, SEM*; Bentler & Dudgeon, 1996), para verificar la dimensionalidad de los ítems. La función estadística elegida corresponden con la metodología para variables categóricas, específicamente será el de máxima verosimilitud con escalamiento ($SB-\chi^2$; Satorra & Bentler, 1994) para reducir el sesgo de no normalidad de los ítems (Tong & Bentler, 2013). El programa que se utilizará será EQS 6.2 (Bentler & Wu, 2012). El ajuste será evaluado mediante indicadores prácticos recomendados por la literatura metodológica (Jackson, Gillaspay & Purc-Stephenson, 2009): *Comparative Fit Index* ($CFI \geq 0.95$), *Standardized Root Mean Square Residual* ($SRMR \leq 0.08$) y *Root Mean Square Error of Approximation* ($RMSEA \leq 0.05$). Alguna modificación post hoc será evaluada examinando los *índices de modificación* (IM; Sorbom, 1989) o *multiplicadores de Langrange* (Bentler, 1986), y la intensidad de su efecto sobre el cambio del modelo; para esto último, se aplicó el método de Saris, Satorra y van der Veld (2009), en que se a priori

especifica un valor mínimo (no trivial), para el parámetro de interés (e.g., carga factorial, error correlacionado, etc.) y una potencia estadística de .80. Como contrapunto teórico, los posibles ajustes al modelo también serán examinado por su relevancia conceptual, para reducir la dependencia de las decisiones en resultados estadísticos que también pueden estar influenciados por variaciones idiosincrásicas de las muestras de participantes recolectadas (Byrne, 2008).

Una de las primeras decisiones en aplicar el modelamiento SEM para validación de instrumentos es el estimador estadístico elegido para evaluar el ajuste del modelo de medición propuesto. Este estimador sirve para producir un estadístico de tipo Chi cuadrado, que proviene de la comparación entre el modelo propuesto, y el modelo estimado estadísticamente. Algún estimador requiere el cumplimiento de algún grado de presunciones respecto a la normalidad multivariada de las variables observables, así como la forma de registro de la variable observada, esto es, si es continua o categórica. El estimador apropiado cuando los datos cumplen la normalidad multivariada es de *máxima verosimilitud* (*maximum likelihood*; Savalei, 2014), pero en caso contrario, existen otros estimadores más precisos y aplicables en el análisis de variables categóricas, como en los ítems ordinales. Por este motivo, para modelar la relación entre los ítems y la variable latente (constructo), se utilizó un estimador para variables categóricas, específicamente *weighted least square mean and variance adjusted* (WLSMV, Asparouhov & Muthén, 2006, 2010). Este método usa un vínculo probit para definir la relación funcional entre los ítems y el constructo, así como correlaciones policóricas entre los ítems, y la estimación de umbrales (*thresholds*) entre categorías de respuestas. Este método permite derivar parámetros (por ejemplo, cargas factoriales) más precisos cuando el número de opciones de

respuesta es menor que 5 y la asimetría distribucional es fuerte (Li, 2016a, 2016b; Sass, Schmitt & Marsh, 2014).

Invarianza

Invarianza de medición. Las comparaciones entre grupos de México y Perú, o intragrupos fueron realizadas en dos fases; la primera será evaluar la *invarianza de medición* de los constructos implicados en el estudio, y para ello se identificarán los grupos potencialmente relevantes que puedan producir variabilidad en las propiedades psicométricas y decidir la existencia del sesgo en las mediciones (Meredith, 1993). Estos dos grupos principales son el género y la nacionalidad de los trabajadores, pues ambas categorías tienden a producir variabilidad en el nivel de intensidad de algunos constructos (Karlqvist, Tornqvist, Hagberg, Hagman, & Toomingas, 2002; Purvanova & Muros, 2010), pero también en factores relacionados con el sesgo de respuesta (Fisher et al., 2009; Harzing, 2006). La nominación de “invarianza de medición” generalmente ocurre en el contexto del análisis del modelamiento de ecuaciones estructurales (Bentler & Dugeon, 1996), y está asociada a la detección del funcionamiento diferencial de ítems (*differential ítem functioning*, DIF; Raju, Laffitte & Byrne, 2002; Teresi, Ramirez, Lai & Silver, 2008; Whitmore & Schumacker, 1999), un término asociado a la teoría de respuesta al ítem (*ítem response theory*, IRT; Teresi et al., 2008). El DIF es la denominación usual para designar a la metodología de evaluación del sesgo dentro del marco de la IRT. Un instrumento compuesto por ítems invariantes o sin DIF, significa que las diferencias en los puntajes pueden ser interpretados como asociadas al constructo, y en menor medida a diferencias en las propiedades psicométricas (Teresi et al., 2008; te Nijenhuis & van der Flier, 2009).

En la aplicación de la invarianza de medición mediante el marco SEM, esencialmente, se aplicará un procedimiento incremental de modelos restringidos

(Meredith, 1993) en que se verifica de manera acumulativa la invarianza *configuracional* (número de dimensiones), *métrica* (cargas factoriales), *invarianza de interceptos* e *invarianza estricta* (residuales). Los criterios para evaluar el grado de invarianza serán evaluados mediante las recomendaciones al respecto (Cheung & Rensvold, 2002), que consiste en la comparación de modelos, específicamente el más restringido frente al menos restringido. Se aplicará la rutinaria prueba estadística $\Delta\chi^2$, y su expresión en índices de ajuste (CFI, RMSEA, SRMR); e indicadores de la magnitud de las diferencias de cada parámetro evaluado por invarianza (Nye & Drasgow, 2011; Nye, Bradburn, Olenick, Bialko, & Drasgow, 2019).

Asociación con otras variables

Como parte de las evidencias de validez, la asociación con otras variables será examinada univariadamente, por medio de asociaciones de monotonía y relaciones lineales. Ambas proveen de información sobre dos formas en que pueden expresarse las relaciones entre variables, es decir, lineal y monotónica no lineal (Chen, & Popovich, 2002). Este análisis será guiado por hipótesis de validez, es decir, que las relaciones esperadas de los constructos medidos por el PROPSIT serán planteadas a priori, y serán evaluadas de acuerdo al rechazo de la hipótesis nula de ausencia de asociación (i.e., $r = 0$), y al tamaño de esta asociación.

3.5.4 Aspectos éticos

Las consideraciones éticas del presente estudio se enfocaron esencialmente en la construcción y recolección de datos; de este modo, para proteger a los participantes que serán evaluados, el proyecto fue evaluado por el Comité de Ética de investigación del CITPSI, con el registro CONBIOÉTICA-17-CEI-003-20190509. En segundo lugar, se analizaron y establecieron las siguientes condiciones:

Confidencialidad. El nombre completo y otro identificador oficial de los participantes no fueron solicitados; por lo tanto, no hay forma de reconocer la identidad del respondiente. La identificación de los participantes (ID), para propósitos de ordenamiento de la base datos, se hizo asignando números consecutivos asociado a condición de la recolección de datos. Este ID no fue asociado a la identidad única del examinado.

Captura de datos. Durante la recolección cuantitativa de datos, la interacción entre el investigador y los participantes ocurrió de manera electrónica. En la primera situación, el investigador no estuvo presente físicamente durante la respuesta del participante, porque existió una interfaz textual en forma de formulario, que contiene las instrucciones y contenido para cumplimentar el PROPSIT, junto con los instrumentos adicionales. Esta forma de recolección eliminó algún impacto directo del investigador sobre reacciones no intentadas del participante.

En la encuesta semiestructurada, se mantuvieron los mismos principios y estrategias como en las encuestas cerradas, es decir, la identidad del entrevistado también fue mantenida en anonimato, mediante la asignación de un código alfanumérico.

3. Resultados

Para el propósito de exposición de resultados, se hizo una presentación modular; esto es, los resultados correspondientes a cada objetivo fueron organizados como un manuscrito empírico y único, con todas sus secciones habituales: introducción, método, resultados, discusión o conclusiones, y referencias.

Todos los resultados fueron organizados en las siguientes partes: La primera parte contiene los resultados de la evaluación del contenido del PROPSIT – FPST para deducir la adecuabilidad de los constructos medidos (Objetivo 1a: Identificar la estructura conceptual de las exposiciones (factores psicosociales); la segunda parte, los resultados de la evaluación de la relevancia, claridad y validez externa de los ítems (Objetivo 1b). Ambos resolvieron la validez del contenido del PROPSIT – FPST.

La tercera parte contiene la identificación de sesgos de respuesta (Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST).

La cuarta parte contiene las evidencias de estructura dimensional (Objetivo 3a); la quinta parte contiene las evidencias de la validez convergente y divergente (Objetivo 3b); la sexta parte contiene las evidencias de la equivalencia de medición (Objetivo 3c); la séptima parte, las evidencias de confiabilidad por consistencia interna (Objetivo 3d). Desde la cuarta a la séptima parte, las evidencias ahí reportadas resuelven la estructura interna del PROPSIT – FPST.

La octava y última parte consistió en presentar los resultados correlacionales entre los puntajes del PROPSIT – FPST y otras variables de efecto (Objetivo 4), aspecto correspondiente con las evidencias de asociación con variables externas.

3.1. Objetivo 1a: Identificar la estructura conceptual de los factores psicosociales del trabajo (exposiciones)

3.1.1. Introducción

El marco teórico para el presente proyecto de validación fue el de *Procesos Psicosociales del Trabajo* (Juárez, 2017; Juárez & Camacho, 2011). Los constructos que la integran están fuertemente influenciados por un marco general para describir y explicar los factores psicosociales del trabajo, específicamente el Modelo Demandas-Recursos (Bakker & Demerouti, 2007; Demerouti, Bakker, Nachreiner & Schaufeli, 2001). Este modelo tiene características que resaltan frente a otros modelos, como la visión heurística de los estresores y sus efectos aplicados al contexto laboral, la inclusión de resultados y procesos de tipo positivo y negativo entre sus componentes, la vinculación entre áreas de aplicación de salud ocupacional y recursos humanos, y la incorporación de los recursos individuales del trabajador (Demerouti, & Bakker, 2011; Schaufeli & Taris, 2014). En el marco del PROPSIT, se identifican varios subsistemas organizativos de los aspectos intra-laborales y extra-laborales (Juárez-García, & Flores-Jiménez, 2020). El foco del presente estudio fue el primer subsistema, el de los factores psicosociales del trabajo de la PROPSIT, para adaptar su contenido al contexto del trabajo peruano. En este marco, los factores psicosociales del trabajo se separan en dos áreas: los factores de riesgo psicosocial (PROPSIT – FRPS), y los factores de recursos favorables (PROPSIT – FRF). Los factores de riesgo psicosocial describen los aspectos del trabajo que interacción con el trabajador para producir malestar, estrés o tensión psicológica, mientras que los factores de recursos favorables son aspectos laborales que promueven la salud, la motivación y el entusiasmo laboral (Juárez, 2017; Juárez & Camacho, 2011).

Bajo este modelo, la construcción teórica del PROPSIT en general, y del PROPSIT – FPST en particular, se desarrolló dentro de un enfoque de investigación emic – etic para hacerlos apropiados para la cultura mexicana (Camacho, 2019, p. 68), pero alineados también con los aspectos psicosociales del trabajo que la literatura internacional consistentemente ha identificado alrededor del mundo. De acuerdo a las investigaciones previas en México sobre los aspectos del trabajo asociados al bienestar y malestar de los trabajadores (Flores & Juárez, 2016; Juárez, 2007; Juárez & Andrade, 2004; Márquez & Juárez, 2018), así como estudios de diseño mixto que sirvieron para verificar y explorar la emergencia de posibles nuevos factores psicosociales emergentes (Camacho, 2019), las dimensiones en el PROPSIT – FRPS que fueron identificadas, y validadas empíricamente (Juárez-García, & Flores-Jiménez, 2020) fueron cuatro: Demandas Laborales, Acoso Laboral psicológico y Liderazgo Estresante; mientras que las dimensiones PROPSIT-FRF identificadas fueron cuatro: Recompensas y Desarrollo Laboral, Control Laboral, Recursos Laborales y Clima de Apoyo Social Laboral.

El objetivo del presente estudio fue explorar los aspectos laborales en una muestra de trabajadores peruanos, para examinar su convergencia con las categorías de contenido y dimensiones del PROPSIT – FRPS, así como la eventual emergencia de nuevos contenidos que podrían utilizarse para construir nuevos factores psicosociales del trabajo. Este objetivo se alinea con el enfoque emic – etic del presente proyecto, en que el contenido identificado en las respuestas de los participantes refleja la variedad de experiencias en los factores psicosociales del trabajo, pero también se examina su vínculo con un marco de referencia. En esta investigación, el marco de referencia fue el PROPSIT – FRPS. De este modo, las respuestas de los participantes fueron codificadas dentro de un análisis que fue principalmente híbrido (Boyatzis, 1998; Fereday & Muir-Cochrane, 2006), en que se inicia con un enfoque

deductivo para la identificación de temas vinculados con el marco teórico de los factores psicosociales del PROPIST, pero permite también la emergencia de posibles contenidos significativos desde un marco inductivo. Los conceptos emergentes o los conceptos corroborados pueden ser descritos narrativa, gráfica o cuantitativamente; particularmente estas dos últimas opciones, gráfica y cuantitativa de reporte añaden claridad y facilidad para comprender los resultados (Pope, Ziebland & Mays, 2000), por ejemplo, mediante la prevalencia resultante de las categorías (Pope et al., 2000) y complementariamente, con la presentación de figuras, como los *word clouds* (Bletzer, 2015; McNaught, & Lam, 2010) o nube de palabras.

Para que el análisis de la codificación e interpretación sea focalizado en líneas de contenido concretas y consistentes (Stuckey, 2015), se creó la siguiente pregunta de investigación: ¿Las situaciones laborales vinculadas con la experiencia de estrés, y con la satisfacción del trabajador en su puesto de trabajo, están alineadas con el PROPSIT?

3.1.2. Método

Participantes

48 trabajadores contactados mediante muestreo no probabilístico, específicamente por oportunidad y accesibilidad con el investigador. Todos residieron en diferentes regiones del Perú. Los trabajadores que aceptaron participar fueron de diferente profesión y ocupación laboral, todos con mayoría legal de edad (> 18 años); los criterios de exclusión fueron no estar trabajando en los últimos seis meses, y haber marcado en el formulario de consentimiento informado, su falta de disponibilidad para participar en el estudio. La descripción cuantitativa de los participantes se muestra en la Tabla 1.

Tabla 1.

Información descriptiva de los participantes

	N	%
Sexo		
Hombre	29	60.4
Mujer	19	39.6
Estado civil		
Soltero	37	77.1
Casado	5	10.4
Conviviente	5	10.4
Viudo	1	2.1
Lugar de nacimiento		
Lima	33	68.8
Fuera de Lima	14	29.2
Sin dato	1	2.1
Grado de instrucción		
Secundaria	8	16.7
Técnico (<3 años)	4	8.3
Técnico (3 años)	19	39.6
Universitario	17	35.4
	M (Md)	DE (Min, Max)
Antigüedad laboral	36.51 (12)	49.3 (3, 24)
Edad	26.5 (25.0)	7.4 (18, 56)
Horas semanales	49.3 (48)	15.0 (7, 84)

Instrumentos

Cuestionario con preguntas abiertas. El estudio fue centrado en las respuestas escritas, y, por lo tanto, con base en la técnica mixta de exploración psicosocial (Camacho, 2016; Flores, & Juárez, 2016; Juárez-García, Flores-Jiménez, & Pelcastre-Villafuerte, 2020; Juárez-García, Flores, & Hindrichs, 2019). Pero operativamente es diferente porque aquí el a) proceso fue completamente descriptivo y no asociativo, b) la codificación no fue abierta sino establecida a priori, debido a la orientación esencialmente emic (i.e., los temas fueron a priori, y basados en el contenido de las dimensiones de factores psicosociales del

trabajo del PROSIT). Por otro lado, comparado con los tamaños muestrales de Camacho (2016; n = 458), Flores, & Juárez (2016; n = 159) y Juárez-García et al. (2020; n = 116), c) el presente estudio fue realizado en muestra pequeña, lo cual reduce la potencia estadística y la capacidad de estabilidad en la generalización de las categorías (desde un ángulo ético y cuantitativo); d) el tamaño muestral se reduce aún más para los análisis adicionales, e.g., los análisis de asociación en Flores y Juárez (2016), o Juárez-García et al. (2020), el tamaño muestral para las asociaciones estadísticas fueron menores de 30, alrededor de 5 veces menos el tamaño muestral inicial); y e) las instrucciones para generar las respuestas a las situaciones laborales negativas fueron más descriptivas.

De este modo, se crearon 2 inductores, cada una con 2 preguntas abiertas se enfocaron en las situaciones laborales como estimuladores de las situaciones estresantes y las situaciones que más le gustan al participante sobre tu trabajo, respectivamente. Se esperó que estas preguntas abiertas facilitarían a) la comprensión de la experiencia del participante de respecto a los factores psicosociales del trabajo, y b) la identificación de temas predeterminados (a priori).

El primer inductor, indagó sobre los factores de riesgo psicosocial, con las siguientes instrucciones: *“Cuando las personas trabajan, ocurren estresores laborales que surgen de manera frecuente o infrecuente. Estresores laborales son situaciones, problemas o eventos que ocurren durante el trabajo de las personas, dificultan trabajar bien y producen incomodidad o malestar, porque uno no puede resolverlos. Para conocer y comprender los estresores que ocurren en su trabajo, por favor, en los siguientes espacios escriba los 3 estresores que le provocan mayor malestar o incomodidad, y que ocurren cuando trabaja; al lado de cada una marque la frecuencia con que ocurren.”* En el siguiente espacio abajo, dentro de una tabla, se ubicaron dos encabezados: “Estresores

laborales (escriba abajo)” (debajo de este encabezado aparecieron tres espacios para las respuestas), y “¿Con qué frecuencia ocurren?” (debajo de este encabezado se presentaron tres opciones: pocas veces, frecuentemente, muchas veces).

En otra tabla, y debajo de la primera, para la indagación de los factores de recursos favorables, se puso la siguiente instrucción: “*En el siguiente recuadro, enumere los 3 aspectos de su trabajo que más le gustan o entusiasman, y también marque la frecuencia con que ocurren*”. Esta instrucción fue breve, comparada con el inductor de las situaciones negativas (que contenían la palabra “estrés”) debido que no contuvo algún concepto que confundirse con otras experiencias, y por lo tanto, no requirió de alguna explicación adicional. Y debajo de esta instrucción, dentro de una tabla idéntica a la anterior, aparecieron los siguientes encabezados una al lado de la otra: “Aspectos que más le gustan o entusiasman de su trabajo (escriba abajo)” (debajo de este encabezado aparecieron tres espacios para las respuestas), y “¿Con qué frecuencia ocurren?” (debajo de este encabezado se presentaron tres opciones: pocas veces, frecuentemente, muchas veces).

Procedimiento

Recolección. Dos colaboradores residentes en Lima participaron en la recolección de datos, aplicando una estrategia de bola de nieve, iniciando la semilla entre los contactos de los evaluadores. Los trabajadores que aceptaron participar respondieron a las preguntas abiertas del cuestionario, junto con las preguntas demográficas incluidas. Esto se hizo en un formulario ad hoc puesto en una plataforma web (cuyo link fue enviado por mensajería WhatsApp), o mediante un archivo de texto MS Word enviado los emails personales. Luego de terminado la recolección y los análisis, los correos electrónicos personales fueron borrados, excepto de los que solicitaron información de los resultados.

Análisis

El proceso de codificación fue esencialmente deductivo, debido a la existencia a priori de un marco temático derivado de las dimensiones y áreas de contenido del PROPSIT. Siguiendo al proceso de codificación de datos resumida por Fereday y Muir-Cochrane (2006), se elaboró el siguiente flujo del proceso de codificación:

1. Desarrollo del manual de códigos. Se hizo un diccionario de códigos derivado apriorísticamente, proveniente de la estructura de factores psicosociales empíricamente validada del PROPSIT en 1339 trabajadores mexicanos (Juárez-García, & Flores-Jiménez, 2020, Tabla 1). Este marco el cual contenía siete factores psicosociales, divididos en factores de riesgo psicosocial (PROPSIT-FRPS: Demandas Laborales, Acoso Laboral psicológico y Liderazgo Estresante), y factores de recursos favorables (PROPSIT-FRF: Recompensas y Desarrollo Laboral, Control Laboral, Recursos Laborales y Clima de Apoyo Social Laboral). Este diccionario contuvo varios elementos (Boyatzis, 1998; Fereday & Muir-Cochrane, 2006) definiciones de los códigos en términos de una *etiqueta* (nombre de las dimensiones específicas), una *definición* (proveniente de la Tabla 1 de Juárez-García & Flores-Jiménez, 2020), y una *descripción* (el contenido de los ítems). Esto se puede ver en el Apéndice.

Este diccionario fue usado como una plantilla – guía para el codificador, desde el cual el codificador lo examinaba cuando quería verificar el código asignado al contenido analizado. La confiabilidad de este diccionario se hizo

2. Codificación. Se elaboró un flujo de cinco pasos para el proceso de codificación mediante la plantilla:

Primero se hizo un escaneo del contenido, que consistió en leer cada respuesta del examinado para familiarizarse con la extensión de las respuestas, y la variedad de las mismas. Esta etapa convergió con un estudio previo (Camacho, 2016).

Segundo, cada respuesta fue leída para ser ubicada en uno de las categorías de contenido existentes; esta ubicación pudo iniciar la categoría general o la categoría específica, y dependió de la rapidez con que el contenido pudo ser identificado en alguna de estas categorías. El proceso de lectura fue guiado por la racionalidad del contenido establecida por dos preguntas-guía que el codificador se hizo a sí mismo: ¿qué significa?, y ¿esto puede ser un ejemplo de qué categoría?

Tercero, el paso 2 interaccionó con la percepción de rapidez para ser identificado en las categorías existentes; de este modo, el calificador asignó el contenido (i.e., respuesta escrita del participante) como *inmediata* (puntaje 3; cuando no se percibió dudas sobre la clasificación y no requirió adicional examen del contenido), *moderadamente iterativa* (puntaje 2; cuando se percibió moderada duda sobre la clasificación inicial, y 3), *intensamente iterativa* (puntaje 1, cuando se percibió alta duda sobre la adecuada clasificación existente, y se requirió ser considerado como no clasificable en las categorías existentes). El puntaje 1 de eficiencia llevaría a los codificadores a crear de una nueva categoría, elaborados independientemente por ellos mismos, y luego revisados por ambos para evaluar la consistencia de su contenido (esto es descrito en el siguiente paso, Cuarto). El procedimiento de la eficiencia de la codificación fue incluido como un medio para fortalecer la credibilidad del estudio (Fereday & Muir-Cochrane, 2006). Si el contenido era ubicado en la categoría específica, la categorización terminaba; si el contenido fue ubicado en la categoría general, se continuó con decidir a qué categoría específica puede ubicarse.

Cuarto, en la situación que en ninguna categoría específica pudo ser utilizada para identificar el contenido analizado, se emergía una categoría específica nueva mediante codificación abierta. Esta nueva categoría fue inducida del contenido específico mediante la creación de una etiqueta que sea genérica, descriptiva en términos generales, y que pueda

resumir y contener a las respuestas específicas de los participantes. Esta categoría emergente fue iterativamente evaluada luego de cada nueva respuesta analizada, de tal modo que la categoría emergente se consolidaba o era modificada para representar mejor el contenido. Las categorías nuevas fueron representadas gráficamente, mediante el *wordcloud* (nube de palabras); se utilizó el programa R *wordcloud* (Fellows, 2018).

3. Calidad del proceso. Para asegurar la legitimidad del proceso de codificación (en la Codificación), se elaboraron dos procedimientos: a) se creó un ejemplo procesual, en que se describía resumidamente los pasos de codificación. Este se entregó al codificador antes de iniciar su tarea, y que le serviría de guía. Por otro lado, b) la confiabilidad del uso del código/plantilla, fue evaluada mediante la repetición de los pasos 1, 2 y 3 por parte de una colaboradora (maestría en enfermería, orientada a investigación y docencia, con experiencia en evaluaciones para validez de contenido, y enfocada en temas distintos a los factores psicosociales del trabajo). El grado de acuerdo entre esta colaboradora y el investigador del proyecto (CAMS) se hizo mediante el coeficiente AC1 (Gwet, 2008). Como fue descrito en el paso 2 del proceso de Codificación, c) la eficiencia de la codificación fue obtenida mediante una calificación del 1 al 3, efectuada por los codificadores cuando asignaron un código al contenido evaluado.

Finalmente, d) la evaluación de la confiabilidad de las respuestas fue evaluada mediante la similaridad de frecuencia con que los estresores y motivadores laborales ocurrieron en la codificación de las tres preguntas abiertas. Para ello se asignaron rankings, iniciando con 1 en la categoría con la más alta frecuencia de aparición dentro de cada pregunta abierta. Estos rankings fueron correlacionados para obtener una medida de similaridad mediante la correlación monotónica de Spearman. Se usaron las funciones *rank*

y *cor* del programa R *base* (R Core Team (2021), y *iccCAC* (Gwet, 2019). En la Figura 8 aparece una descripción gráfica del proceso.

4. Reporte de resultados. Los resultados incluirán la presentación de los indicadores de calidad del proceso, y también la prevalencia o frecuencia de los estresores y motivadores en las categorías elegidas. Esta cuantificación de la prevalencia permitió llevar la interpretación de los resultados a un ángulo objetivo, y con mayor oportunidad de ser comprendido (Pope, et al., 2000). También se construyeron gráficos basados en nubes de palabras (*word clouds*), para visualizar la prevalencia y fortalecer la claridad de los resultados (Bletzer, 2015; McNaught, & Lam, 2010). Cada palabra graficada estuvo directamente vinculada con su prevalencia tabulada, de tal modo que el tamaño de la palabra dibujada indicó la fuerza de su frecuencia. Las diferentes palabras, aun las que fueron parecidas conceptualmente iguales o similares, no fueron homogeneizadas en una palabra común, para evitar alterar la variabilidad de las palabras ocurrida.

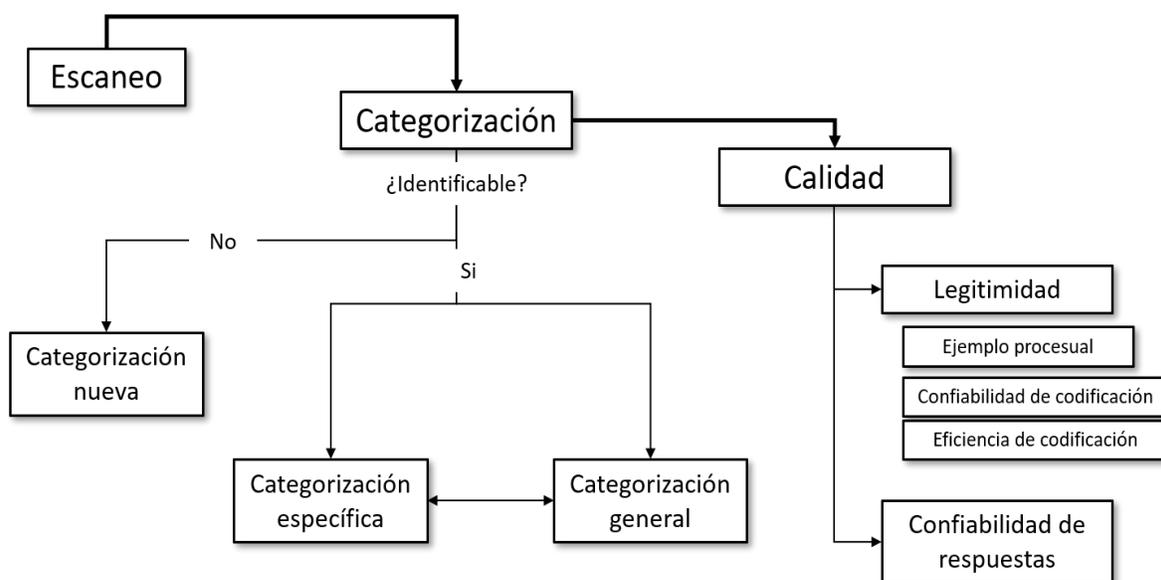


Figura 8

Visualización del proceso de codificación

3.1.3. Resultados

Eficiencia de la codificación

La eficiencia de la codificación de los estresores (Tabla 2) se incrementó de la primera respuesta (estresor 1) a la tercera respuesta (estresor 3). Esto también fue indicado por la asociación monotónica, debido que el tamaño de esta asociación se incrementó en esta misma secuencia. Por otro lado, respecto a los motivadores laborales, la eficiencia fue predominantemente en el valor 3 (codificación inmediata), y la asociación ordinal entre los motivadores laborales fue $|1.0|$ (Tabla 3). El coeficiente gamma entre el motivador 1 y motivador 2 fue negativa debido a un solo valor discrepante (el Motivador1 “concesionaria” fue asignado el valor 1, y consecuentemente una nueva categoría fue creada). Sin este valor en todos, la igualdad absoluta fue detectada entre ambos motivadores ($gamma = 1.0$).

Tabla 2

Eficiencia de la codificación de los estresores

	Estresor 1		Estresor 2		Estresor 3	
	N	%	N	%	N	%
Frecuencia						
3 Inmediata	3	6.3	36	75.0	38	79.2
2 Iterativa moderada	26	54.2	8	16.7	5	10.4
1 Iterativa intensa	17	35.4	1	2.1	-	-
Sin respuesta	2	4.2	3	6.3	5	10.4
Asociación monotónica ^a						
Estresor 1	-	-	-	-	-	-
Estresor 2	.26 ^{ns}		-	-	-	-
Estresor 3	.47 ^{ns}		.53*		-	-

Nota. ^aLas asociaciones son coeficientes *gamma*. ns: estadísticamente no significativo. * $p < .05$

Tabla 3

Eficiencia de la codificación de los motivadores laborales

	Motivador 1		Motivador 2		Motivador 3	
	N	%	N	%	N	%
Frecuencia						
3 Inmediata	46	95.8	44	91.7	44	91.7
2 Iterativa moderada	0	0	2	4.2	2	4.2
1 Iterativa intensa	1	2.1	0	.0	0	.0
Sin respuesta	1	2.1	2	4.2	2	4.2
Asociación ordinal ^a						
Motivador 1	-	-	-	-	-	-
Motivador 2	1.0 ^b		-	-	-	-
Motivador 3	1.0 ^b		1.0		-	-

Nota. ^a Coeficientes *gamma*. ^b Coeficientes gamma sin un dato discrepante.

Confiabilidad de la codificación

El colaborador codificó 20 respuestas elegidas aleatoriamente, obtenidas de la primera pregunta abierta de los estresores laborales y de los motivadores laborales (para ambos, también fueron elegidos aleatoriamente). Se obtuvieron coeficientes AC1 de .82 ($p < .05$; IC 95% = .63, 1.00, e.e. = .09) y .38 ($p < .05$; IC 95% = .11, .65, e.e. = .12) respectivamente, indicando que el proceso de codificación mantuvo un nivel adecuado de acuerdo entre-codificadores y replicabilidad en los estresores, pero de magnitud baja en los

motivadores laborales. En este último, sin embargo, los temas específicos (e.g., apoyo de compañeros, apoyo de jefes y clima de unión) se interpretan dentro de una sola dimensión general del PROPSIT (e.g., clima y apoyo social en el trabajo), y en conjunto el efecto de este bajo acuerdo puede considerarse no significativa para la interpretación del PROPSIT en base a puntajes.

Categorías en factores de riesgo psicosocial

En la Tabla 4, se presentan la frecuencia y porcentaje de aparición de cada código, así como rankings de frecuencia (1 asignado a la categoría con mayor frecuencia), para las tres respuestas (i.e., estresor 1, estresor 2 y estresor 3). Para resumir los resultados con mejor claridad, se reorganizó esta distribución en la Tabla 5, de acuerdo a tres grupos reconocibles: categorías frecuentes ($\geq 10.0\%$), más o menos frecuentes ($> .0\%$) e infrecuentes ($.0\%$). Desde esta Tabla 5, se observan temas muy recurrentes (e.g., carga y ritmo de trabajo). En general, más de la mitad de las categorías fueron aplicadas a las respuestas de los participantes en cada estresor, indicando la relevancia de estas categorías para los participantes muestreados. Las categorías infrecuentes, donde el porcentaje fue $.0\%$ fue predominante en el contenido referido al acoso por los superiores o por los compañeros.

Las respuestas inclasificables fueron asignadas a categorías nuevas, inferidas de la lectura y reflexión de cada respuesta. En la Figura 9, se muestra el *wordcloud* las respuestas a cada estresor, así como categorías emergidas. Estas categorías emergentes, debido a su frecuencia, principalmente la percepción del pobre trabajo en equipo (e.g., “incumplimiento de los compañeros”, “desorganización”, etc.), clima conflictivo (“clima laboral no agradable”, “problemas de comunicación”), y temas genéricos o sin detalles (e.g., “preocupación”, “ambiente laboral”, “problemas laborales”). Otras respuestas pueden considerarse también factores psicosociales (e.g., “falta de materiales”, “pagos

impuntuales”, etc.), mientras otras se refieren a efectos (“preocupación”). Las dos primeras categorías emergentes parecen referirse a eventos involucrados en las relaciones interpersonales orientadas hacia la tarea y hacia la mantención de la sociabilidad positiva.

Como nota adicional, la asociación monotónica entre los rankings de frecuencia de los estresores (estresores con mayor frecuencia fueron asignados a 1, como aparecen en el sub-encabezado R, Tabla 4) fue alta ($\geq .80$; Tabla 8), indicando que los mismos estresores aparecieron consistentemente en las tres preguntas efectuadas.

Tabla 4

Prevalencia de las categorías en factores de riesgo psicosocial

Categorías	Estresor 1 (n = 46)			Estresor 2 (n = 45)			Estresor 3 (n = 44)		
	N	%	R	N	%	R	N	%	R
Carga y ritmo de trabajo	14	30.4	2	8	16.7	2	9	20.5	2
Alta responsabilidad y peligrosidad	1	2.2	7	4	8.3	5	2	4.5	6
Jornadas, turnos u horarios	4	8.7	3	3	6.3	6	6	13.6	3
Cognitivas o atencionales	0	0.0	9	0	0.0	10	2	4.5	6
Emocionales (trato con personas)	4	8.7	3	5	10.4	3	2	4.5	6
Esfuerzo físico	0	0.0	9	1	2.1	9	0	0.0	10
Ambiente físico	4	8.7	3	5	10.4	3	5	11.4	4
Acoso Por superiores	0	0.0	9	0	0.0	10	0	0.0	10
Acoso Por compañeros o subordinados	0	0.0	9	0	0.0	10	0	0.0	10
Supervisión controlante	4	8.7	3	2	4.2	7	3	6.8	5
Retroalimentación negativa/ inadecuada del supervisor	1	2.2	7	2	4.2	7	2	4.5	6
No clasificable	15	32.6	1	15	31.3	1	12	27.3	1

Nota. N: número de respuestas analizadas en cada factor. %: porcentaje de N. R: ranking asociado a N (1 asignado al mayor N, y así consecutivamente).

Tabla 5

Prevalencia de la clasificación en factores de riesgo psicosocial

Distribución	Estresor 1 (n = 46)	Estresor 2 (n = 45)	Estresor 3 (n = 44)
Frecuentes ($\geq 10.0\%$)	<ul style="list-style-type: none"> • Carga y ritmo de trabajo, • No clasificable 	<ul style="list-style-type: none"> • Carga y ritmo de trabajo, • Emocionales (trato con personas), • Ambiente físico, • No clasificable 	<ul style="list-style-type: none"> • Carga y ritmo de trabajo, • Jornadas, turnos u horarios, • Ambiente físico, • No clasificable
Más o menos frecuentes ($>0.0\%$)	<ul style="list-style-type: none"> • Alta responsabilidad y peligrosidad, • Jornadas, turnos u horarios, • Emocionales (trato con personas), • Ambiente físico, • Supervisión controlante, • Retroalimentación negativa/ inadecuada del supervisor 	<ul style="list-style-type: none"> • Alta responsabilidad y peligrosidad, • Jornadas, turnos u horarios, • Esfuerzo físico, • Ambiente físico, • Supervisión controlante, • Retroalimentación negativa/ inadecuada del supervisor 	<ul style="list-style-type: none"> • Alta responsabilidad y peligrosidad, • Cognitivas o atencionales, • Emocionales (trato con personas), • Supervisión controlante, • Retroalimentación negativa/ inadecuada del supervisor
Infrecuentes (0.0%)	<ul style="list-style-type: none"> • Cognitivas o atencionales, • Esfuerzo físico, • Acoso Por superiores • Acoso Por compañeros 	<ul style="list-style-type: none"> • Cognitivas o atencionales, • Acoso Por superiores, • Acoso Por compañeros 	<ul style="list-style-type: none"> • Esfuerzo físico, • Acoso Por superiores • Acoso Por compañeros

Nota. La presente distribución se deriva de la Tabla 4.

Categorías en factores de recursos favorables

En la Tabla 6 se presentan la frecuencia y porcentaje de aparición de cada categoría de los factores de recursos favorables, así como rankings de frecuencia (1 asignado a la categoría con mayor frecuencia), para las tres respuestas (i.e., motivador 1, motivador 2 y motivador 3). Para resumir los resultados con mejor claridad, se reorganizó esta distribución en la Tabla 7, de acuerdo a tres grupos reconocibles: categorías frecuentes ($\geq 10.0\%$), más o menos frecuentes ($> .0\%$) e infrecuentes ($.0\%$). Es claro la predominancia de la percepción de lo gratificante de las tareas y el clima de unión, que se encuentran entre las más consistentes, junto con contenidos no clasificables inicialmente. Por otro lado, las categorías consistentemente infrecuentes fueron la percepción de justicia organizacional, la seguridad en la conservación del empleo, el trabajo variado, y la claridad de funciones y

roles. Las categorías emergentes (Figura 10), derivadas de las respuestas no clasificadas inicialmente, parecieron definir temas como la recreación y la limpieza/orden con predominancia.

Como nota adicional, la asociación monotónica entre los rankings de frecuencia de los motivadores (estresores con mayor frecuencia fueron asignados a 1, como aparecen en el sub-encabezado R, Tabla 6) varió moderadamente ($.83 \leq rho \leq .42$; Tabla 8), indicando que los mismos estresores no aparecieron consistentemente en las tres preguntas efectuadas.

Tabla 6

Prevalencia de la clasificación en factores de recursos favorables

Categorías	Motivador 1 (n = 47)			Motivador 2 (n = 46)			Motivador 3 (n = 46)		
	N	%	R	N	%	R	N	%	R
Justicia organizacional	0	0.0%	9	0	0.0%	13	0	0.0%	11
Salario motivante	2	4.3%	7	3	6.5%	6	3	6.3%	6
Reconocimiento por el trabajo	2	4.3%	7	1	2.2%	9	7	14.6%	2
Posibilidades de desarrollo laboral y profesional (estatus)	0	0.0%	9	3	6.5%	6	2	4.2%	8
Seguridad en la conservación del empleo	0	0.0%	9	0	0.0%	13	0	0.0%	11
Tarea gratificante y con trascendencia	10	21.3%	2	6	13.0%	3	5	10.4%	4
Influencia-autonomía en el trabajo	4	8.5%	6	5	10.9%	4	1	2.1%	9
Utilización de habilidades en el trabajo	0	0.0%	9	1	2.2%	9	1	2.1%	9
Trabajo variado (no monótono)	0	0.0%	9	0	0.0%	13	0	0.0%	11
Funciones y roles claros	0	0.0%	9	1	2.2%	9	0	0.0%	11
Recursos materiales, equipos, herramientas	0	0.0%	9	1	2.2%	9	0	0.0%	11
Capacitación, adiestramiento	5	10.6%	4	2	4.3%	8	3	6.3%	6
Apoyo de compañeros	5	10.6%	4	9	19.6%	1	0	0.0%	11
Apoyo de jefes	0	0.0%	9	0	0.0%	13	4	8.3%	5
Clima de unión	13	27.7%	1	8	17.4%	2	7	14.6%	2
No clasificable	6	12.8%	3	5	10.4%	4	9	18.8%	1

Nota. N: número de respuestas. %: porcentaje de N. R: ranking asociado a N.

Tabla 7
Redistribución de la clasificación en factores de recursos favorables

Distribución	Motivador 1 (n = 47)	Motivador 2 (n = 46)	Motivador 3 (n = 46)
Frecuentes ($\geq 10.0\%$)	<ul style="list-style-type: none"> Tarea gratificante y con trascendencia, Capacitación, adiestramiento, Apoyo de compañeros, Clima de unión, No clasificable 	<ul style="list-style-type: none"> Tarea gratificante y con trascendencia, Influencia-autonomía en el trabajo, Apoyo de compañeros, Clima de unión, No clasificable 	<ul style="list-style-type: none"> Reconocimiento por el trabajo, Tarea gratificante y con trascendencia, Clima de unión No clasificable
Más o menos frecuentes ($>0.0\%$)	<ul style="list-style-type: none"> Salario motivante, Reconocimiento por el trabajo, Influencia-autonomía en el trabajo 	<ul style="list-style-type: none"> Salario motivante, Reconocimiento por el trabajo, Desarrollo laboral y profesional (estatus), Utilización de habilidades en el trabajo, Funciones y roles claros, Recursos materiales, Capacitación, adiestramiento. 	<ul style="list-style-type: none"> Salario motivante, Desarrollo laboral, Influencia-autonomía en el trabajo, Utilización de habilidades en el trabajo, Capacitación, adiestramiento, Apoyo de jefes
Infrecuentes (0.0%)	<ul style="list-style-type: none"> Justicia organizacional, Desarrollo laboral y profesional (estatus), Seguridad en la conservación del empleo, Utilización de habilidades en el trabajo, Trabajo variado (no monótono), Funciones y roles claros, Recursos materiales, Apoyo de jefes 	<ul style="list-style-type: none"> Justicia organizacional, Seguridad en la conservación del empleo, Trabajo variado (no monótono), Apoyo de jefes 	<ul style="list-style-type: none"> Justicia organizacional, Seguridad en la conservación del empleo, Trabajo variado (no monótono), Funciones y roles claros, Recursos materiales, Apoyo de compañeros

Nota. La presente distribución se deriva de la Tabla 6.

Tabla 8

Asociación entre los rankings (Rnk) de estresores y motivadores laborales

Estresores			
	Rnk estresor 1	Rnk estresor 2	Rnk estresor 3
Rnk estresor 1	1		
Rnk estresor 2	.91	1	
Rnk estresor 3	.92	.80	1

Motivadores			
	Rnk motivador 1	Rnk motivador 2	Rnk motivador 3
Rnk motivador 1	1		
Rnk motivador 2	.83	1	
Rnk motivador 3	.63	.42	1

Nota. Rnk: ranking del motivador o estresor

Figura 9

Wordcloud de respuestas a los tres inductores de estrés y categorías emergentes

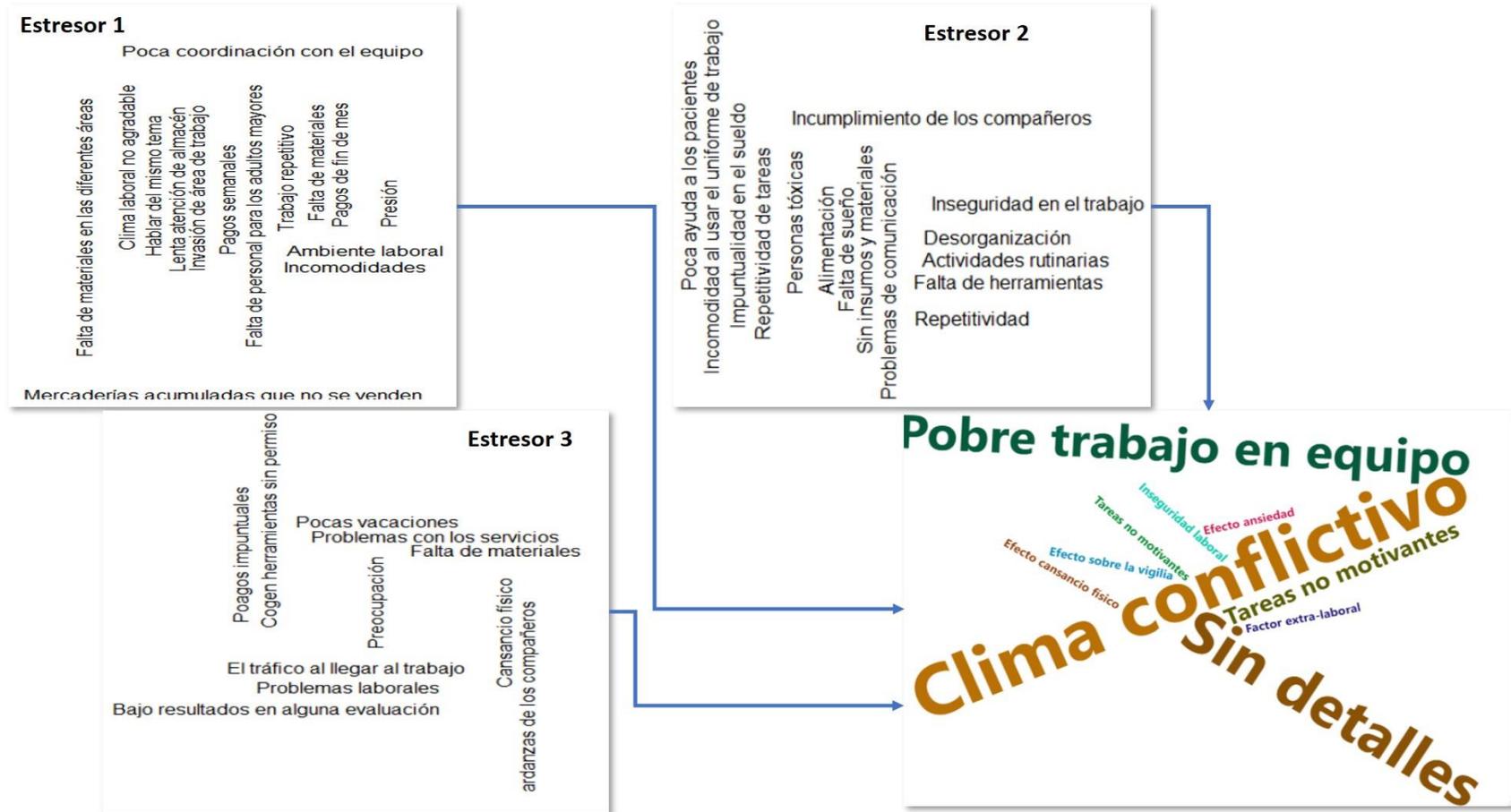
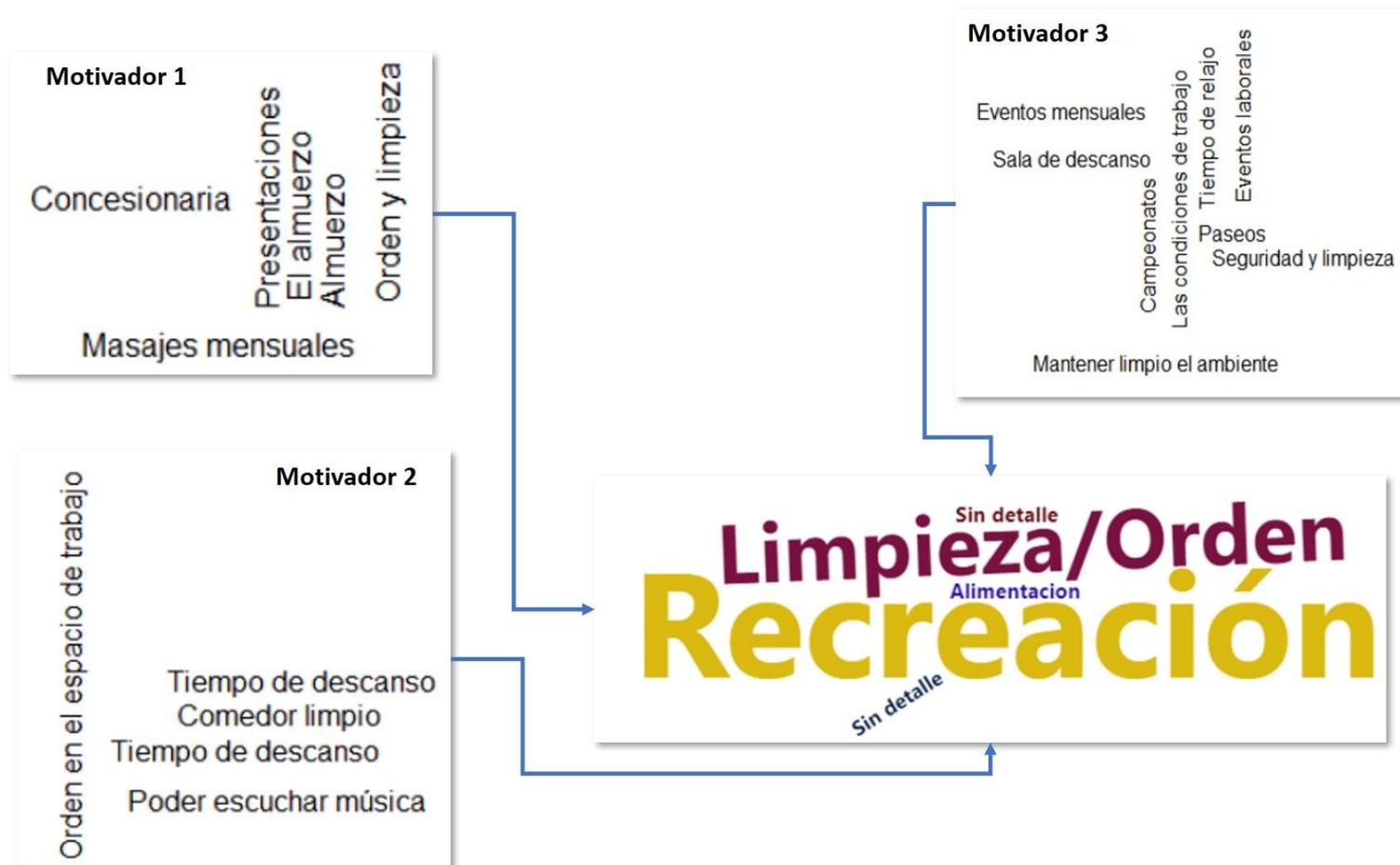


Figura 10

Wordcloud de Respuestas a los tres inductores de motivación y categorías emergentes



3.1.4. Discusión

Respecto a la evaluación de la calidad del proceso, debido que la revisión en general fue iterativa y secuencial, la eficiencia de la codificación mejoró sustancialmente del primer estresor al tercer estresor, debido que la inmediata codificación (puntaje 3) aumentó de 6.3% en el primer estresor, hacia 77% aproximadamente. Esto indica que las codificaciones en los dos últimos estresores fueron percibidas como clasificaciones más seguras, y fácilmente reconocidas cuando se asignaron los códigos a las respuestas abiertas de los participantes. Esto también se expresó en la asociación ordinal, debido que la asociación fue moderadamente fuerte entre la eficiencia de los dos últimos estresores. Este patrón mejora en la eficiencia fue repetido en la calificación de los motivadores laborales, comparado con los estresores, porque excepto algunas muy pocas valoraciones, la codificación de todos los motivadores fueron inmediatamente hechos.

En otra parte de la evaluación de la calidad del proceso, en la evaluación de la similaridad de frecuencia con que los estresores y motivadores laborales ocurrieron en la codificación, se halló alta similaridad en la ocurrencia de las categorías en las tres preguntas abiertas ($\rho > .79$), sugiriendo que los estresores identificados fueron replicables en estas tres oportunidades en que los trabajadores respondieron. Una implicación de esto que las categorías fueron replicables, y que representan un patrón constante en diferentes trabajos y ocupaciones. Sin embargo, respecto a los motivadores laborales, la similaridad fue moderada, y disminuyó en las dos últimas respuestas. Esto implica que el contenido de los recursos favorables es menos replicable, y que hay factores posiblemente nuevos para el grupo de trabajadores peruanos muestreados.

Respecto a las categorías de las dimensiones de factores psicosociales y recursos favorables, éstas fueron predominantemente eficaces para capturar los contenidos de las respuestas de los participantes. Estas respuestas fueron generadas tres veces, y, por un lado, sirvieron para evaluar la reproducibilidad de las respuestas, y por otro, para maximizar la variabilidad de respuesta. Las categorías en los factores de riesgo psicosocial fueron más consistentes, y sugiere el valor emic de estas categorías para representar contenidos relevantes de los constructos latentes. Por otro lado, en los factores de recursos favorables, los contenidos específicos mostraron moderada consistencia, y pueden representar las experiencias de la muestra peruana de manera eficaz. En esta área también se detectó que varias categorías fueron de frecuencia cero, y la cantidad de categorías no aparecidas puede dar la impresión sobre una sobreabundancia de categorías irrelevantes en la muestra de trabajadores peruanos. Sin embargo, este resultado puede ser infraestimado porque cada categoría estuvo representada por un solo ítem del PROPSIT, y la oportunidad que esta categoría parezca conceptualmente irrelevante fue maximizada.

Una situación concreta que debe ser resaltada fue el contenido del acoso ejercido por los superiores o compañeros fue inexistente, así como algunas otras (e.g., demandas cognitivas y de esfuerzo físico); éstas no parecen relevantes para la experiencia de los trabajadores peruanos muestreados porque la frecuencia de aparición fue cero. En algunas regiones puede ser tan alarmante alto (Hong Kong: entre 39.1% y 58.9%; Ng, & Chan, 2021), pero también relativamente bajo (Malasia: 11.2%; Awai, Ganasegeran, & Abdul Manaf, 2021). En Perú, en el contexto laboral de médicos residentes, se reportó una alta prevalencia (61.1%; Nieto-Gutierrez et al., 2018), lo que parece contrastar con la ausencia de este estresor en la muestra peruana. La ausencia de este contenido puede sugerir que la particularidad de la muestra participante pudo influir en el subreporte de tipo de factor de

riesgo psicosocial, pero no necesariamente que este factor sea inexistente. Esto también puede aplicarse en los otros contenidos absolutamente infrecuentes, y es probable que se vincule con la restricción del muestreo de experiencias en una muestra de menos de 50 trabajadores, o que la muestra desde el cual se generaron las respuestas en los estudios mexicanos fuera principalmente homogénea. Por ejemplo, empleados de empresa de manufactura (Camacho, 2016), comerciantes informales (Juárez-García et al., 2020) e instructores comunitarios (Flores, & Juárez, 2016).

Una implicación de las categorías nuevas es que éstas representan potenciales dimensiones no muestreadas en el PROPSIT-FSPS, y que pueden conducir a crear ítems que definan un contenido apropiado al contexto peruano, y una mejor cobertura de los constructos de factores psicosociales del trabajo. La emergencia de estas nuevas categorías, sin embargo, no es definitiva para crear nuevas dimensiones, porque requiere mayor amplitud de respuestas y muestreo de trabajadores. Como se anotó en las líneas anteriores, en contraste con los estudios mexicanos en que se estudió la emergencia estresores o “motivadores” laborales (Camacho, 2016; Flores, & Juárez, 2016; Juárez-García et al., 2020), el muestreo del presente estudio se hizo en un tamaño muestral inferior y con amplitud de puestos de trabajos y carreras. Por lo tanto, la definición de los temas emergentes en los factores de riesgos psicosocial y los factores de recursos favorables pueden ser considerados hipótesis de trabajo para exploraciones con mayor potencial estadística, y mayor oportunidad de obtener respuestas consistentes.

Respecto a las limitaciones, primero el tamaño de la muestra participantes es una aparente restricción para obtener mayor variedad de respuestas, debido que con más participantes hay una mayor oportunidad de más respuestas, y posiblemente todas las categorías mostrarían algún nivel de frecuencia, con exclusión de la infrecuencia absoluta

(ver resultados en la Tabla 5 y 7). Segundo, debido a restricciones en la disponibilidad de los participantes, así como en la proximidad del investigador con ellos, no se hizo una exploración adicional de las respuestas, especialmente de aquellas inicialmente clasificadas como no clasificables en referencia las categorías existentes. Una entrevista aplicada sobre estos contenidos inicialmente no clasificables puede ser requerido, como en la técnica mixta original (Camacho, 2019; Juárez, Flores, & Imke, 2019).

Una vez explorados los constructos desde el modelo PROPSIT, los contenidos que serán presentados a los participantes en formato de preguntas de encuesta, pueden ser evaluados respecto a su claridad. Esta evaluación forma parte de las indagaciones de validez de contenido, y fueron resueltas en la siguiente parte.

3.2. Objetivo 1b: Validez de contenido: Claridad y relevancia de los ítems del PROPSIT

El presente objetivo resolvió tres aspectos en la evaluación de la validez de contenido orientado hacia la claridad de los ítems, y específicamente en la construcción de medidas de factores psicosociales: a) la participación de examinados participantes, b) la legibilidad o lecturabilidad de los ítems, y c) el consenso o acuerdo estadístico. La claridad se refiere al grado en que los ítems son entendibles, claros, o comprensibles para una persona. Esta propiedad condiciona al avance de la construcción o adaptación de una medida, debido que no se pueden obtener evidencias de relevancia conceptual si el contenido primero no es claro.

¿Participantes vs jueces expertos?

La evaluación de la validez de contenido tiene dos enfoques procedimentales generales, uno cualitativo y otro cuantitativo. En cualquiera de ambos aspectos, los aspectos de contenido que habitualmente se evalúan son la relevancia y la claridad. La claridad y la relevancia son categorías generales, pero existen otros aspectos que también pueden ser evaluados, como la utilidad, la importancia (Zamanzadeh et al., 2015), la familiaridad con el contenido (Grant, J. S., & Davis, L. L. (1997)), la ubicación factorial (Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003), la sensibilidad para diferenciar el estatus del examinado (Guillot-Valdés, Guillén-Riquelme, & Buela-Casal, 2022), ofensividad (Guillot-Valdés et al., 2022), impacto emocional (Merino-Soto, 2018), etc. Estos dependen del uso del instrumento, y pueden ser aspectos de contenido secundarios, porque un contenido sin relevancia conceptual ni claridad, no podrían ser evaluados en aspectos condicionados a ambos.

Para la evaluación de algunos de estos aspectos, habitualmente se obtienen impresiones de participantes que sirven como evaluadores del grado en que el contenido del instrumento cumple con tales aspectos. Estos participantes pueden ser las personas a quienes se destina el instrumento o expertos de contenido (i.e., “jueces expertos”, *subject matter experts*). El diseño frecuente de un estudio de validez de contenido incluye a jueces expertos, pero no es raro incluir también a participantes.

Existen unas posibles desventajas cuando se consideran a participantes, como el grado de instrucción (i.e., la calidad y cantidad de años estudios no es constancia y puede ser una fuente de varianza de los resultados), las competencias de lenguaje (i.e., *literacy* y habilidades funcionales), o la falta de familiaridad frente a las tareas de validez de contenido (Edwards, Thomas, Rosenfeld, & Booth-Kewley, 1997; Sheatsley, 1983). Especialmente, las competencias de lenguaje puede ser un efecto moderador siempre presente en la obtención de juicios sobre la claridad del contenido de los ítems. Esto es, la claridad intrínseca del contenido del ítem puede interaccionar con la competencia de lenguaje del participante. Por otro lado, hay ventajas importantes: los participantes tienen directa relación con el contenido (e.g., pueden formar parte de la muestra – objetivo del estudio), su experiencia puede considerarse realista (i.e., no imaginada, simulada o vicaria), y la variedad de experiencias asociadas puede ser mayor en cantidad y similaridad (e.g., comparados con los jueces expertos).

Los jueces expertos también llevan limitaciones no ignorables (Edwards et al., 1997): la habilidad y competencias de lenguaje son más altas en el investigador y su equipo, comparadas con los examinados; pueden ocurrir creencias incorrectas sobre la capacidad de los participantes para comprender el contenido de los ítems (Sheatsley, 1983). Estas creencias incorrectas pueden estar motivadas por sesgos cognitivos perceptuales, como la familiaridad percibida, la experticia autopercebida, etc.; también, puede ocurrir inconsistencia entre la lógica probabilística (i.e., la relación

causal probabilística entre eventos) y lógica determinística (i.e., la relación causal única entre eventos), en una manera en que los jueces expertos consideran su juicio como determinista. En un terreno práctico, la falta de disponibilidad para aceptar o para cumplimentar la tarea es una experiencia común entre investigadores.

Por lo tanto, el presente estudio enfatiza la participación de participantes del estudio, particularmente en las etapas pre-piloto o piloto de la creación o adaptación de medidas. Esta participación requiere corroborar la magnitud de *literacy* o competencia general del lenguaje.

Lectorabilidad

Existen otros potenciales componentes de la validez de contenido que parecen ser menos atendidos en los estudios psicométricos en general, y específicamente en las medidas de FPST. Uno de estos es la lectorabilidad de un texto, y se refiere al grado de comprensibilidad por parte del lector (i.e., facilidad de comprensión, etc.; Edwards et al., 1997; Ríos, 2017). Esta propiedad se conduce mediante una aproximación cuantitativa del grado en que un texto es leíble, y es particularmente útil en el proceso de construcción textual de los ítems de una medida, pero como herramienta complementaria (Edwards et al., 1997). Es frecuentemente utilizada en la evaluación de materiales textuales en ciencias de la salud (Goodman, Griffey, Carpenter, Blanchard, & Kaphingst, 2015; Ríos, 2017; Wang, Miller, Schmitt, & Wen, 2013), y en ciencias de la educación (Parker, Hasbrouck, & Weaver, 2001).

Dado su vínculo conceptual con la claridad de los ítems (Goodman et al., 2013), la evaluación de la lectorabilidad es otro camino para evaluar los ítems en contraste con la evaluación basada en el juicio percibido de los jueces expertos o de los participantes. La evaluación de la lectorabilidad, sin embargo, tiene algunas limitaciones: a) su formulación matemática es variable, específicamente en los componentes de legibilidad del texto considerados importantes, así como la ponderación de estos componentes en la fórmula; b) existen fórmulas para idioma español e inglés, no pueden usarse de manera intercambiable; c) algunas fórmulas son simples y consistentes (e.g., *Simple Measure of*

Gobbledygook, *SMOG*; Wang et al., 2013), pero solo son aplicados a textos anglosajones; d) los niveles de interpretación de la lecturabilidad obtenidas de éstas fórmulas pueden no ser consistentes entre ellas (Parker et al., 2001); y e) tienden a ser débiles predictores de lenguaje (Parker et al., 2001). Las limitaciones anteriores indican que su uso debe ser acompañado con otras evaluaciones, para de tal modo lograr convergencia. Más aún cuando existen muchas fórmulas y modificaciones de los niveles de lecturabilidad.

Consenso estadístico

Una medida sumaria de tendencia central requiere que la variabilidad de las unidades no sea fuertemente dispersa. Esta fuerte variabilidad prácticamente limita la validez de la media u otra medida de tendencia. Por lo tanto, la variabilidad es una precondition para interpretar un resultado como los coeficientes para estimar la cuantía de validez de contenido. Por ejemplo, el coeficiente V (Aiken, 1980, 1985) está algebraicamente formulado como la distancia de la media de las calificaciones de jueces respecto a una propiedad como relevancia, importancia, claridad u otros aspectos de contenido. Este coeficiente perdería validez si los datos tuvieran una distribución uniforme, o polarizada. Un ítem no puede ser evaluado mediante este coeficiente, a menos que remueva la fuente de fuerte variación.

El presente objetivo propone una metodología envolvente de tres pasos para evaluar rigurosamente la validez de contenido enfocado en la claridad de los ítems. Esto consiste en a) evaluar la lecturabilidad de los ítems mediante diferentes fórmulas aplicables a habla hispana, b) evaluar las competencias de lenguaje (literacy) para identificar una posible covariable controlable estadísticamente, c) estimar el consenso de las evaluaciones de claridad de los ítems.

Método

3.2.1. Sección 1: Análisis de la lecturabilidad del contenido de PROPSIT - FPST

Participantes (unidades analizadas). Las unidades analizadas fueron los ítems de la evaluación de factores psicosociales del trabajo (PROPSIT – FPST), que contiene 41 ítems.

Instrumento (Legibilidad). La herramienta fue una interfase visual en plataforma web, escrito en español y ubicado en <https://legible.es/>. Esta interface contiene una caja de texto, instrucciones de uso, y botones de acción para iniciar el análisis.

Procedimiento.

Aplicación. El contenido factores psicosociales del trabajo fue evaluado en dos conjuntos: todos los ítems en una sola estimación, y cada ítem independientemente. En el análisis de cada ítem, se omitió el punto final y la numeración.

Análisis. Se usó el software *analizador de legibilidad* (Muñoz, 2020), contenido en la plataforma web <https://legible.es/>. El análisis incluyó varios coeficientes de lecturabilidad en lugar de un solo. Se buscó explorar la convergencia entre los diferentes coeficientes mediante un enfoque de consenso entre coeficientes, en que el nivel de lecturabilidad predominante fue evaluada por el autor del presente proyecto. Los índices de lecturabilidad utilizados fueron: el índice de Fernández Huerta (1959), el índice de perspecuidad (Szigrist – Pazos, 1993), la escala INFLESZ (Barrio, 2008) y el índice μ (mu), debido que están adaptados para textos de habla hispana.

Resultados de la Lecturabilidad

Lectorabilidad de los 41 ítems. Se hallaron los resultados que aparecen en la Tabla 9, en que el nivel aproximado de lecturabilidad fue el recomendado para un adulto joven, recién egresado de estudios secundarios, pero aún sin estudios universitarios, o un poco más. La convergencia entre los métodos usados no fue difícil de establecer, debido que todos convergieron, excepto el índice μ (mu).

Tabla 9

Lectorabilidad del PROPSIT – FPST, usando todos los 41 ítems

Índice	Valor	Dificultad
FH (Fernández Huerta)	63.26	Normal
SP (Sziegriszt-Pasos)	58.22	Normal
IN (INFLESZ)	58.22	Normal
Mu	48.57	Difícil

Lectorabilidad ítem por ítem. Respecto a la lectorabilidad ítem por ítem, se halló una distribución normal en las estimaciones para los ítems, en los cuatro coeficientes (Figura 11).

De acuerdo al coeficiente Fernández Huerta (FH), los ítems se encontraron en un grado de lectorabilidad entre algo fácil y algo difícil; según el coeficiente Sziegriszt-Pasos (SP) entre bastante fácil y bastante difícil; de acuerdo a INFLESZ (IN) entre algo fácil y bastante difícil, y según el coeficiente μ (mu) entre un poco fácil y un poco difícil. Por otro lado, hubo moderada convergencia en los ítems detectados como difíciles o muy difíciles de leer (ver Tabla 10).

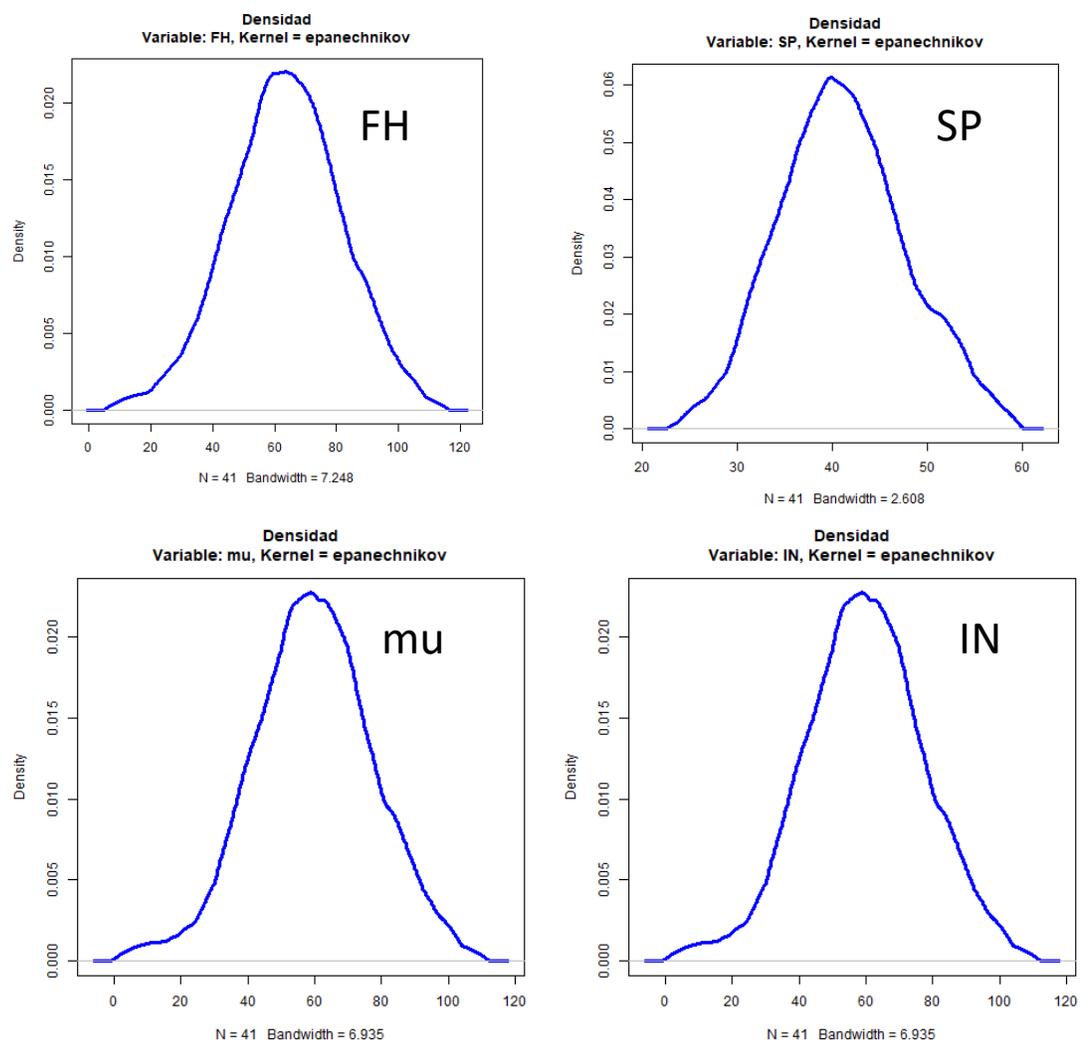


Figura 11

Distribución de densidad kernel de los coeficientes de lectorabilidad en cada ítem

Tabla 10

Ítems considerados difíciles de acuerdo a los cuatro coeficientes de lecturabilidad

Fernández – Huerta (FH)		Szigriszt-Pazos (SP)		INFLESZ (IN)		Mu	
Nivel	Ítem	Nivel	Ítem	Nivel	Ítem	Nivel	Ítem
Muy difícil	36	Muy difícil	--	Muy difícil	36, 22, 11, 26, 32	Muy difícil	36
Difícil	29, 22, 11, 26, 32, 17, 19	Árido	11, 36, 22, 32, 26, 27, 35, 19, 18, 17	-	-	Difícil	22, 11, 26, 32, 17, 19, 18, 37, 35, 27

Nota. El contenido completo de los ítems aparece en el Apéndice.

Conclusión. Se concluye que la lecturabilidad del contenido total del PROPSIT – FPS puede considerarse apropiada para el nivel mínimo educación alcanzada por trabajadores, es decir, secundaria nivel básico. Por otro lado, en el nivel de los ítems, éstos varían entre un rango de facilidad y moderada dificultad de lecturabilidad. Sin embargo, este resultado es aproximado debido a la limitada extensión del texto analizado en cada ítem. En los ítems donde hubo convergencia moderada en identificar ítem difíciles, esto implica la revisión de los ítems para ser reformulados en su redacción y puedan ser más lecturables o fáciles de comprender. Sin embargo, la aparente limitada claridad de estos ítems requiere ser corroborada con la percepción de claridad derivado de los participantes.

3.2.2. Sección 2: Análisis de las competencias del lenguaje de los participantes (literacy)

Participantes. 54 trabajadores contactados no probabilísticamente. Todos residieron en diferentes regiones del Perú. Los trabajadores que aceptaron participar fueron de diferente profesión y ocupación laboral, todos con mayoría legal de edad (> 18 años); los criterios de exclusión fueron no estar trabajando en los últimos seis meses, y no haber respondido algún ítem en las medidas aplicadas de evaluación de la competencia del lenguaje (i.e., SALI y SILS; ver el siguiente párrafo). Se distribuyeron en 27 varones (50.0%) y 26 mujeres (48.1%), y uno sin identificación (1.9%), entre 18 y

56 años ($M = 26.5$, $DE = 7.57$). las carreras se distribuyeron en: administración y negocios internacionales (3, 5.6%), ciencias de la comunicación (1, 1.9%), contabilidad (3, 5.6%), ingenierías (8, 14.8%), mecánica automotriz y de mantenimiento (9, 16.7%), psicología (6, 11.1%), docencia (1, 1.9%), enfermería (1, 1.9%), y carreras técnicas (modistería, técnico en enfermería, cosmetología, panificación, electricidad, farmacia, secretariado ejecutivo, mecánico; 12, 22.2%); 10 (18.5%) no se identificaron.

Instrumentos. Fueron aplicados tres instrumentos:

Formulario de claridad de los ítems. Diseñado para obtener las calificaciones de claridad de los ítems, por parte de los participantes. Se usó una escala ordinal de cinco puntos, desde *nada claro* hasta *completamente claro*. Con este formulario, cada examinado determinó, en una escala de respuesta del 1 (nada claro o entendible) al 5 (completamente claro o entendible), el grado en que cada ítem es percibido como claro y entendible.

The Self-Assessed Literacy Index (SALI; Olson et al., 2011). Cuestionario de auto-reporte consistente en dos partes, competencia evaluada y práctica de competencias. La primera contiene una instrucción raíz (“*Considerando el idioma castellano que tú manejas, ¿qué tan bueno eres...*”) aplicada a cuatro ítems sobre la capacidad de comprensión oral, expresarse oralmente, leer y escribir. A más puntaje obtenido de la suma de sus ítems, más competencia de lenguaje auto-evaluada. Se responden con cuatro opciones de respuesta. En el Apéndice se muestra una copia del instrumento. El SALI fue independientemente traducido por el autor, y por las dos colaboradoras de la recolección de datos; y en conjunto evaluaron las traducciones hasta alcanzar un consenso.

Single Item Literacy Screen (SILS; Goodman et al., 2015). Consistió en una adaptación del ítem original cuyo contenido fue “*¿Con qué frecuencia tiene problemas para conocer sobre su situación médica, debido a dificultades para entender la información escrita?*”. El contenido adaptado para su aplicación general fue “*¿Con qué frecuencia tiene problemas para comprender información o*

instrucciones escritas? La respuesta es expresada en una escala ordinal de cinco puntos (desde *siempre* hasta *nunca*). A mayor puntuación obtenida de la suma de los ítems, menos competencia de lenguaje auto-evaluada. En el Apéndice se muestra una copia del instrumento. Para interpretar los resultados del SILS consistente con SALI, la puntuación de SILS fue invertida, de tal modo que las puntuaciones 1, 2 y 3 se interpretan como limitada competencia, mientras 4 y 5 se interpretan como competencia adecuada. Su traducción siguió el mismo procedimiento hecho para SALI.

Procedimiento

Recolección. Dos colaboradores residentes en Lima participaron en la recolección de datos, aplicando una estrategia de bola de nieve, iniciando la semilla entre los contactos de los evaluadores. Los trabajadores que aceptaron participar respondieron primero sobre su auto-evaluación de competencias de lenguaje (literacy), y segundo, evaluaron cada ítem mediante la calificaron la claridad del ítem. Esto se hizo en un formulario ad hoc puesto en una plataforma web (cuyo link fue enviado por mensajería WhatsApp), o mediante un archivo de texto MS Word enviado los emails personales. Luego de terminado la recolección y los análisis, los emails personales fueron borrados, excepto de los que solicitaron información de los resultados.

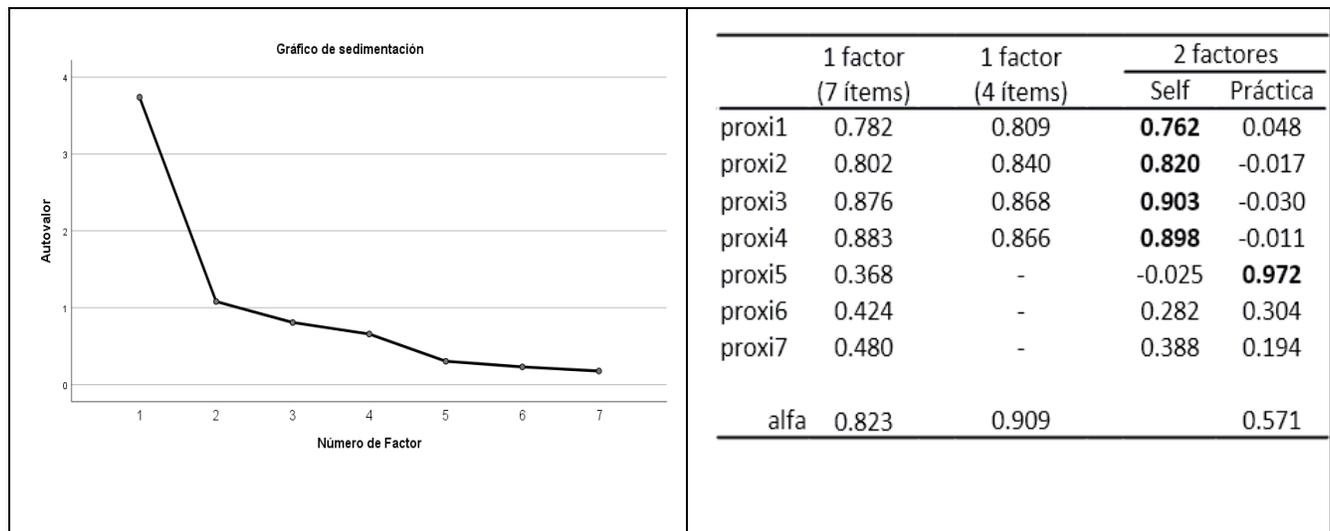
Análisis. Consistió en lo siguiente: a) identificar la dimensionalidad del SALI, mediante un análisis factorial exploratorio (con estimador mínimos cuadrados no ponderados, ULS) sobre las correlaciones policóricas entre los ítems, para obtener el número de dimensiones latentes, la validez de los ítems y la confiabilidad de consistencia interna; b) para determinar la validez de constructo mediante la asociación monótonica (correlación rho de Spearman) de las medidas de competencias de lenguaje (i.e., SALI y SILS) con las calificaciones de claridad de los ítems; c) comparar los grupos de bajo y alta competencia de lenguaje deducidas de SALI y SILS, se usó la diferencia estandarizada (mediante el programa *effsize*; Torchiano, 2020), respecto a las calificaciones de claridad de los ítems;

Resultados. Se presentan tres resultados: dimensionalidad del SALI, correlaciones de validez de constructo de las medidas de competencias de lenguaje con las calificaciones de claridad del ítem, y finalmente la comparación entre grupos de adecuada y limitada competencia.

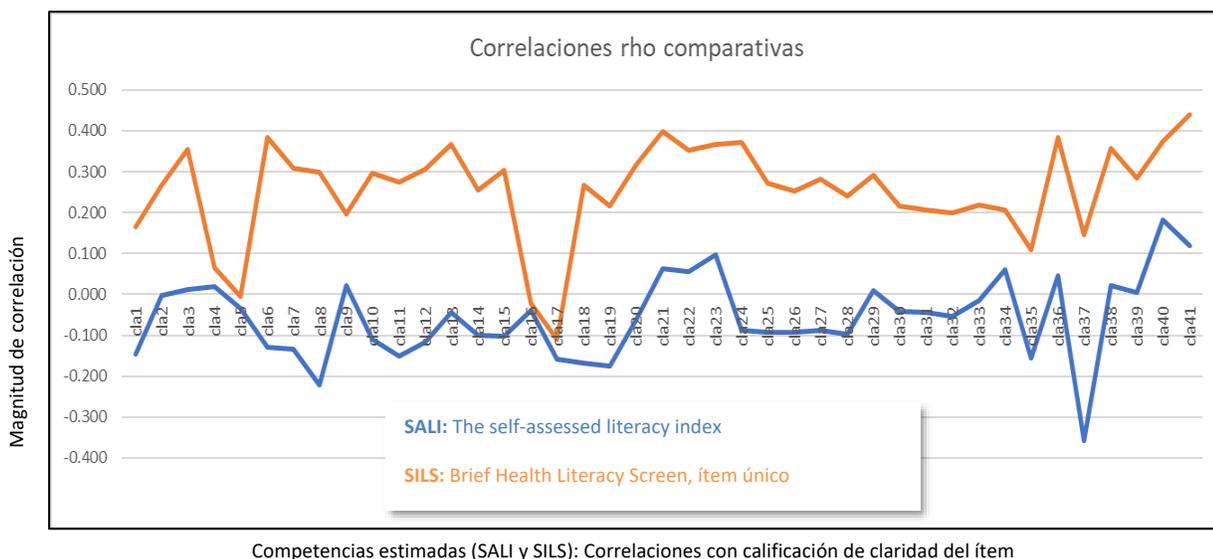
Dimensionalidad del SALI. En la Tabla 11 se muestran los resultados del análisis factorial exploratorio del SALI. Se halló que una sola dimensión define mejor el constructo medido por SALI. Sin la inclusión de los ítems sobre la práctica de competencias (ítems 5, 6 y 7), la unidimensionalidad se define por los cuatro primeros ítems, con altas cargas factoriales. La confiabilidad alfa de este factor es muy alta.

Tabla 11

Dimensionalidad de SALI (análisis factorial exploratorio)



Validez de constructo: Correlación SALI – SILS con calificaciones de claridad. Las correlaciones *rho* Spearman entre SILS y las calificaciones de claridad fueron sistemáticamente más altas, en términos absolutos y comparativos frente a SALI (ver Figura 12). Por otro lado, las correlaciones SALI fueron inconsistentes, porque fueron positivas o negativas; mientras SILS mostró correlaciones siempre positivas con la calificación de claridad del ítem (CCI). Se concluye que SILS es un potencial candidato para diferenciar participantes con limitada o adecuada competencia de lenguaje.



Competencias estimadas (SALI y SILS): Correlaciones con calificación de claridad del ítem

Figura 12

Competencias estimadas (SALI y SILS): Correlaciones con calificación de claridad del ítem

Comparación de grupos adecuada vs baja competencia. Por medio del SILS, 46 participantes fueron seleccionados con adecuadas competencias (puntajes 4 o 5 en el SILS), mientras que 8 participantes fueron identificados como limitados en sus competencias (puntajes 1, 2, o 3 en el SILS). Los resultados (ver Tabla 12) pueden resumirse en lo siguiente: los participantes clasificados como adecuados casi siempre produjeron calificaciones de claridad más altas, el tamaño de la diferencia fue predominantemente sustancial, algunos ítems (en rojo) mostraron patrón opuesto, pero puede representar 10% de variación aleatoria (i.e., 90% de confianza).

Tabla 12

Comparación de grupos de competencia de lenguaje (adecuada vs limitada)

Ítems	Adecuado (4 a 5) n = 46		Limitado (1 a 3) n = 8		<i>d</i>	Adecuado (4 a 5) n = 46		Limitado (1 a 3) n = 8		<i>d</i>	
	M	DE	M	DE		M	DE	M	DE		
Riesgos psicosociales						Recursos psicosociales favorables					
cla1	4.76	.52	4.38	.92	.64	cla23	4.74	.68	3.88	1.55	.99
cla2	4.46	.91	4.13	1.36	.33	cla24	4.74	.80	4.13	1.13	.70
cla3	4.63	.71	4.00	1.41	.73	cla25	4.61	.83	4.63	.52	-.02
cla4	4.63	.80	4.75	.71	-.15	cla26	4.61	.93	4.13	1.36	.47
cla5	4.54	.84	4.50	1.41	.05	cla27	4.50	.96	3.63	1.77	.77
cla6	4.82	.49	3.38	1.77	1.77	cla28	4.80	.54	4.63	.74	.31
cla7	4.61	.91	4.25	1.04	.38	cla29	4.70	.55	4.38	1.06	.49
cla8	4.63	1.00	4.13	1.36	.47	cla30	4.59	.78	4.50	1.07	.10
cla9	4.61	.77	4.50	.76	.14	cla31	4.70	.63	4.63	.52	.11
cla10	4.65	.80	4.00	1.51	.69	cla32	4.78	.42	4.63	.52	.36
cla11	4.70	.59	4.25	1.39	.58	cla33	4.72	.54	4.38	1.41	.46
cla12	4.78	.55	4.13	1.25	.93	cla34	4.74	.49	4.38	1.41	.52
cla13	4.52	.91	3.63	1.69	.83	cla35	4.80	.58	4.50	1.07	.44
cla14	4.65	.80	4.63	.52	.03	cla36	4.78	.66	3.75	1.58	1.19
cla15	4.57	.89	3.75	1.75	.76	cla37	4.85	.47	4.88	.35	-.06
cla16	4.52	.94	4.75	.71	-.25	cla38	4.78	.55	4.00	1.41	1.04
cla17	4.61	.83	5.00	.00	-.50	cla39	4.78	.47	4.25	1.39	.78
cla18	4.54	1.05	3.50	1.85	.86	cla40	4.63	.83	3.88	.99	.87
cla19	4.39	1.26	3.50	2.07	.62	cla41	4.65	.67	3.63	1.19	1.31
cla20	4.46	1.09	3.50	1.69	.79	-	-	-	-	-	-
cla21	4.50	.75	3.38	1.60	1.20	-	-	-	-	-	-
cla22	4.74	.54	4.00	1.31	1.04	-	-	-	-	-	-

Nota. *d*: diferencia estandarizada

Se concluye que: a) las calificaciones de claridad del contenido fueron afectadas con la inclusión de personas con limitada competencia de lenguaje, b) las diferencias entre las personas con limitadas y adecuadas competencias de lenguaje, fueron claras y generalmente sustanciales, c) la estimación de la competencia de lenguaje mediante un solo ítem (SALI) fue efectiva, d) para propósitos de controlar el efecto moderador de la competencia del lenguaje sobre las calificaciones de claridad, los participantes identificados con limitada competencia deben ser excluidos en los análisis siguientes (consenso y claridad).

3.2.3. Sección 3: Análisis del Consenso / acuerdo

Participantes. Del número total de participantes iniciales ($n = 54$), de acuerdo a los resultados de las competencias del lenguaje, se excluyeron 8 trabajadores; la muestra para este análisis se redujo por lo tanto a 46.

Instrumento. Se usó el mismo formulario aplicado en el análisis de las competencias del lenguaje.

Procedimiento. La recolección de datos y el análisis se describen a continuación.

Recolección. El mismo procedimiento anterior fue también aplicado para la recolección de datos (i.e., calificaciones de claridad de los ítems).

Análisis. como una medida de tamizaje cruda de la falta de consenso, se aplicó el sistema AJUS (Galtung, 1969) para identificar cuatro tipos de distribuciones en cada ítem: *A* (distribución unimodal con normalidad), *J* (distribución unimodal con asimetría), *U* (distribución bimodal o respuesta polarizada en los extremos) y *S* (distribución multimodal y/o uniforme). Las distribuciones identificadas como *A* y *J* sugieren consenso, y el resto falta de consenso. Se complementó con un indicador numérico, el coeficiente de dispersión ordinal *ADm* (average deviation from mean; Burke, & Dunlap, 2002) y no condicionado a una distribución nula a priori (Burke, & Dunlap, 2002; Meyer et al., 2014). Menos magnitud de *ADm* sugiere menos dispersión; el punto de corte para identificar fuerte dispersión fue .80 (# de categorías/6; Burke, & Dunlap, 2002). Se usó el programa *agrmt* (Ruedin, 2020).

Resultados. Los resultados se presentan en la Tabla 13. Estos pueden resumirse en lo siguiente: a) las distribuciones de ítems predominantemente codificadas como *J* (sistema Galtung), b) todos los ítems expresaron consenso en la calificación de ítems, c) algunos ítems mostraron posible desacuerdo, pero el análisis distribucional confirmó el consenso, y d) la sensibilidad de criterios pudo producir falsos positivos en identificar falta de consenso.

Se concluye que las calificaciones de claridad de los ítems del PROPSIT – FPST son esencialmente consensuadas. Los participantes pueden producir este tipo de patrones de respuesta, y garantizar posteriores análisis

Tabla 13

Tipo de distribución y dispersión ordinal de los ítems del PROPSIT - FPST

Recursos Psicosociales			Riesgos Psicosociales			
Ítems	AJUS	ADm	Ítems	AJUS	ADm	Anotaciones
cla23	J	.390	cla1	J	.328	-
cla24	J	.385	cla2	J	.697	-
cla25	J	.529	cla3	J	.518	-
cla26	J	.566	cla4	J	.551	-
cla27	J	.674	cla5	J	.582	-
cla28	J	.298	cla6	J	.242	-
cla29	J	.397	cla7	U	.581	Sugiere repuesta polarizada, pero ADm < .80
cla30	J	.551	cla8	U	.575	-
cla31	J	.451	cla9	J	.571	-
cla32	J	.319	cla10	J	.526	-
cla33	J	.405	cla11	J	.447	-
cla34	J	.341	cla12	J	.298	-
cla35	J	.275	cla13	J	.660	-
cla36	J	.344	cla14	J	.468	-
cla37	J	.203	cla15	J	.613	-
cla38	J	.308	cla16	J	.637	-
cla39	J	.287	cla17	J	.543	-
cla40	J	.490	cla18	J	.646	-
cla41	J	.467	cla19	J	.870	Alta dispersión, pero distribución sugiere unimodalidad
-	-	-	cla20	J	.734	Moderada alta dispersión, pero distribución J y ADm < .80
-	-	-	cla21	J	.616	-
-	-	-	cla22	J	.386	-

Nota. J: distribución unimodal asimétrica. U: distribución multimodal. Participantes: 46 luego de aplicar el criterio de competencias de lenguaje.

3.2.4. Sección 4: Análisis de la claridad de los ítems

Participantes. Del número total de participantes iniciales ($n = 54$), de acuerdo a los resultados de las competencias del lenguaje, se excluyeron 8 trabajadores; la muestra para el análisis se redujo por lo tanto a 46.

Instrumento. Se usó el mismo formulario aplicado en el análisis de las competencias del lenguaje, es decir, el formulario de percepción de claridad de los ítems.

Procedimiento. La recolección de datos y el análisis se describen a continuación.

Recolección. El mismo procedimiento anterior fue también aplicado para la recolección de datos (i.e., calificaciones de claridad de los ítems).

Análisis. Se utilizó el coeficiente V (Aiken, 1980) para cuantificar el grado de claridad percibida; éste varía entre 0 y 1, y mientras más cerca de 1.0 con más seguridad se concluye en la validez del contenido. Se implementaron intervalos de confianza asimétricos (Merino & Livia, 2009), en el nivel del 90%. El criterio de aceptabilidad para indicar a un ítem como claro en su contenido fue que el límite inferior del intervalo de confianza debe ser igual o mayor a .60 (Merino-Soto, 2018). Se usó el programa *ICaiken* (Merino & Livia, 2009).

Resultados. En la Tabla 14 se muestran los resultados de la evaluación de la claridad. Se halló que: a) todos los ítems cumplieron con el criterio $V \geq .60$, b) la claridad de los ítems es percibida con gran acuerdo, debido que todos los ítems mostraron $V > .83$. Se concluye que los ítems son percibidos como claros en la manera como fueron redactados, con altos niveles de claridad percibida.

3.2.5. Sección 4: Análisis de la relevancia del contenido de PROPSIT – FPST

Participantes. Fueron 15 participantes con una edad media de 27.6 (min = 23, max = 46), todos trabajadores entre de 1 a 8 años de experiencia laboral, con estudios universitarios distribuidos entre licenciatura (n = 12), maestría (n = 1) y especialización en psicología organizacional/industrial (n = 2). Fueron de un amplio rango de carreras); y excepto un participante, todos fueron empleados por contrato en instituciones privadas o públicas.

Tabla 14

Evaluación de la claridad percibida de los ítems (coeficiente V)

Ítems	V	I.C. 90%		Inf > .60	Ítems	V	I.C. 90%		Inf > .60
		Inf.	Sup.				Inf.	Sup.	
Riesgos psicosociales					Recursos psicosociales				
cla1	.940	.904	.963	Sí	cla22	.935	.898	.959	Sí
cla2	.865	.818	.901	Sí	cla23	.935	.898	.959	Sí
cla3	.908	.866	.937	Sí	cla24	.935	.898	.959	Sí
cla4	.908	.866	.937	Sí	cla25	.903	.860	.933	Sí
cla5	.885	.841	.918	Sí	cla26	.903	.860	.933	Sí
cla6	.955	.923	.974	Sí	cla27	.875	.829	.910	Sí
cla7	.903	.860	.933	Sí	cla28	.950	.916	.971	Sí
cla8	.908	.866	.937	Sí	cla29	.925	.887	.951	Sí
cla9	.903	.860	.933	Sí	cla30	.898	.855	.929	Sí
cla10	.913	.872	.941	Sí	cla31	.925	.887	.951	Sí
cla11	.925	.887	.951	Sí	cla32	.945	.910	.967	Sí
cla12	.945	.910	.967	Sí	cla33	.930	.892	.955	Sí
cla13	.880	.835	.914	Sí	cla34	.935	.898	.959	Sí
cla14	.913	.872	.941	Sí	cla35	.950	.916	.971	Sí
cla15	.893	.849	.925	Sí	cla36	.945	.910	.967	Sí
cla16	.880	.835	.914	Sí	cla37	.962	.932	.980	Sí
cla17	.903	.860	.933	Sí	cla38	.945	.910	.967	Sí
cla18	.885	.841	.918	Sí	cla39	.945	.910	.967	Sí
cla19	.847	.799	.886	Sí	cla40	.908	.866	.937	Sí
cla20	.865	.818	.901	Sí	cla41	.913	.872	.941	Sí
cla21	.875	.829	.910	Sí	-	-	-	-	-
					-	-	-	-	-

Nota. V: coeficiente V. Inf.: límite inferior. Sup: límite superior. Inf. > .60: criterio para establecer la claridad del ítem.

Participantes: 46 luego de aplicar el criterio de competencias de lenguaje. El contenido completo de los ítems aparece en el Apéndice.

Instrumento. Se usó el mismo formulario aplicado en el análisis de las competencias del lenguaje, en que se incluyó el formulario de percepción de claridad de los ítems, y de relevancia conceptual de los ítems. el escalamiento de respuesta fue el mismo de cinco opciones, pero esta vez las etiquetas de cada punto de la escala fueron, del 1 al 5: *nada relevante, poco relevante, aceptablemente relevante, muy relevante, completamente relevante*. Una definición de cada constructo, así como de los contenidos específicos fue presentada para el conjunto de ítems correspondiente.

Procedimiento

Recolección. El mismo procedimiento anterior fue también aplicado para la recolección de datos (i.e., calificaciones de claridad de los ítems). Esto es, los participantes evaluaron la claridad primero, y luego la relevancia conceptual de los ítems, en el mismo formulario.

Análisis. Se utilizó el coeficiente V (Aiken, 1980) para cuantificar el grado de relevancia conceptual percibida del ítem en relación a la definición de su dimensión. V varía entre 0 y 1, y mientras más cerca de 1.0 con más seguridad se concluye en la relevancia del contenido. Se implementaron intervalos de confianza asimétricos (Merino & Livia, 2009), en el nivel del 90%. El criterio para aceptar la relevancia percibida por parte de los jueces fue que el límite inferior del intervalo de confianza debe ser igual o mayor a .60 (Merino-Soto, 2018). Se usó el programa *ICaiken* (Merino & Livia, 2009).

Resultados. En la Tabla 15, se observa que todos los ítems pertenecientes al Factores de Riesgo Psicosocial ($V > .77$) y Factores de Recursos Favorables ($V > .75$) fueron altos y sobre .60. Del mismo modo, el límite inferior del intervalo de confianza de cada coeficiente V fue superior al criterio ($> .68$ y $> .64$, respectivamente), indicando su significancia estadística, y que posible extensión a la población de referencia. De acuerdo a estos resultados, se concluye que

3.2.6. Conclusión general

Los ítems del PROPSIT – FPS, individualmente o en conjunto, mostraron predominante lecturabilidad apropiada para trabajadores que como mínimo tuvieron secundaria básica. Luego de controlar el efecto de la potencial limitada competencia de lenguaje e identificar la falta de consenso entre los participantes, se detectó una alta claridad percibida de los ítems del PROPSIT – FPS, con valores V superiores a .84. El estudio tiene implicaciones metodológicas para evaluar la claridad de contenido, enfatizando la rigurosidad y la reducción de la práctica de usar solo un método para estimar los estadísticos de interés.

Tabla 15

Evaluación de la relevancia conceptual (n = 15)

Ítem	V	Inf	Sup	Ítem	V	Inf	Sup
Factores de Riesgo Psicosocial				Factores de Recursos Favorables			
item1	0.93	0.86	0.97	item22	0.87	0.78	0.92
item2	0.92	0.84	0.96	item23	0.92	0.84	0.96
item3	0.78	0.69	0.86	item24	0.90	0.82	0.95
item4	0.93	0.86	0.97	item25	0.90	0.82	0.95
item5	0.93	0.86	0.97	item26	0.90	0.82	0.95
item6	0.90	0.82	0.95	item27	0.97	0.90	0.99
item7	0.88	0.80	0.94	item28	0.90	0.82	0.95
item8	0.92	0.84	0.96	item29	0.90	0.82	0.95
item9	0.88	0.80	0.94	item30	0.75	0.65	0.83
item10	0.93	0.86	0.97	item31	0.93	0.86	0.97
item11	0.90	0.82	0.95	item32	0.95	0.88	0.98
item12	0.92	0.84	0.96	item33	0.82	0.72	0.89
item13	0.90	0.82	0.95	item34	0.97	0.90	0.99
item14	0.83	0.74	0.90	item35	0.98	0.93	1.00
item15	0.88	0.80	0.94	item36	0.98	0.93	1.00
item16	0.92	0.84	0.96	item37	0.97	0.90	0.99
item17	0.95	0.88	0.98	item38	0.97	0.90	0.99
item18	0.98	0.93	1.00	item39	1.00	0.96	1.00
item19	0.97	0.90	0.99	item40	0.98	0.93	1.00
item20	0.90	0.82	0.95	item41	0.95	0.88	0.98
item21	0.92	0.84	0.96				

Nota. V: coeficiente V. Inf.: límite inferior. Sup: límite superior. Inf. > .60: criterio para establecer la claridad del ítem. El contenido completo de los ítems aparece en el Apéndice.

3.3. Objetivo 1c: Validez de contenido: estructura de respuesta de los ítems y validez con criterios externos

3.3.1. Introducción

En la etapa de creación y/o adaptación de medidas, el desglose de información univariada de variables individuales que conforman los puntajes compuestos, es un contenido que se considera muy importante para comprender mejor el comportamiento cuantitativo de las escalas o test, pero también para propósitos complementarios tales como la replicabilidad, la evaluación meta-analítica, el cumplimiento de supuestos y en general, buenas prácticas de reporte de resultados cuantitativos (Appelbaum et al., 2018; Cooper, 2020). Por extensión de lo anterior y de manera más específica, las propiedades univariadas de los ítems puede ser extremadamente útil para revelar la tendencia de respuesta de los evaluados, la ubicación de los ítems en el rango del puntaje observado, y la variabilidad del atributo en los comportamientos específicos capturados por los ítems. Por ejemplo, en una medida que contiene ítems con diferente fuerza asociativa con criterio de interés, su capacidad predictiva con el criterio podrá también variar. Respecto al atributo latente en estos ítems, el puntaje factorial u observado mostrará consecuentemente un decremento en su potencial predictivo con estos ítems (Urbina, 2014).

La investigación de la validez de los ítems habitualmente se enfoca en la relación ítem-constructo dentro del mismo instrumento, comúnmente llamada formalmente *estructura interna* (AERA et al., 2014). En un instrumento multidimensional, ésta evalúa y describe la asociación del ítem en su propia dimensión y frente a otras dimensiones. Esta fuente de evidencia de validez está alineada con dos propiedades de la estructura interna de un constructo: su *validez convergente* y *validez divergente* (Farrell, & Rudd, 2009; Fornell & Larcker, 1981). Esto es, la fuerza de la varianza de los ítems dentro de su constructo frente a la varianza con los otros factores no intentados para medirlo. Un paso adicional puede hallarse evaluando la escalabilidad de cada ítem, habitualmente representado en

estimaciones basadas en el modelo de Rasch (Boone, & Staver, 2020), donde los ítems y el atributo latente están asociados en la misma escala.

El desarrollo de estos aspectos ayuda a caracterizar la validez de los ítems para representar su constructo dentro del contenido conceptual del instrumento, pero la validez de los ítems puede extenderse a incluir su asociación con criterios externos y otras características estructurales de los ítems, como el funcionamiento de sus opciones de respuesta. Por ejemplo, si los ítems creados para evaluar hostigamiento laboral teóricamente deberían asociarse con síntomas de depresión (e.g., Figueiredo-Ferraz, Gil-Monte, Olivares-Faundez, 2015; Torell et al., 2015) entonces esta covariabilidad debe observarse no solo en la escala total, sino también en estimaciones de dependencia estadística entre los ítems y criterios empíricos de depresión, específicamente identificando el tamaño de estas asociaciones, así como su dirección positiva o negativa, y su significancia estadística. Esta propiedad ítem-criterio externo ha sido conocida como *validez del ítem* (Price, 2017, p. 191; Urbina, 2014, p. 254), sin embargo, es el aspecto que parece menos abordado en la construcción o adaptación de instrumentos en la literatura contemporánea, y específicamente en la investigación de factores psicosociales del trabajo. Efectivamente, esto se corrobora en el reducido tratamiento de este tema en libros de evaluación y psicometría (e.g.: Coulacoglou, & Saklofske, 2017, todo el libro; Morrison & Embretson, 2018, capítulo tres; Price, 2017, capítulo seis; Urbina, 2014, capítulo seis), o en las recomendaciones de buenas prácticas actuales de validación de medidas (e.g., Boateng, Neilands, Frongillo, Melgar-Quíñonez, & Young, 2018). La red de asociaciones de los ítems con otros constructos fuera de su modelo de medición da forma y sentido a lo que se conoce como *red nomológica* (Preckel, & Brunner, 2017), en la cual se hipotetiza, se analiza y verifican las relaciones entre constructos mediante la construcción de hipótesis de relación esperadas. En el nivel de los ítems, esto significa configurar las hipótesis de relación convergente ($r \neq 0$) o divergente ($r \cong 0$) de los ítems con otros constructos. La fuerza de la validez de los ítems en estos aspectos tiene relevancia para la

interpretación del puntaje del instrumento en que los ítems forman parte (Urbina, 2014), porque pone en evidencia el valor teórico y aplicado del contenido de los ítems.

Cuando se aborda el tema de la asociación entre variables, un aspecto que se introduce es la diferenciación entre la dependencia lineal y no lineal. En el estudio de la dependencia bivariada entre variables durante la construcción y adaptación de medidas en ciencias sociales, predomina la exploración de relaciones lineales (Speed, 2011), sin embargo, es razonable que existan otros tipos de asociación. En la detección de asociaciones bivariadas, pueden existir asociaciones lineales, monotónicas y no monotónicas (Fujita et al., 2009), y la investigación psicológica no ha permanecido ajena a la exploración de diferentes asociaciones más allá de la sola dependencia lineal entre variables. En la literatura psicológica sobre la investigación de constructos del desarrollo y en otros contextos aplicados, las relaciones no lineales han aparecido como un modelo de covariabilidad que representa válidamente la asociación entre las variables (Makonnen, 2019). Por ejemplo, en variables tales como el desarrollo de habilidades visomotoras en un rango de edad sensible de cambio (e.g., Merino, 2011), en el cambio de roles, estatus y personalidad en adultos (e.g., Helson, R., & Soto, 2005), y en la expresión de sintomatología depresiva (e.g., Besser, Priel, Flett, & Wiznitzer, 2007). En el contexto laboral, también se ha explorado la relación del estrés con las habilidades de afrontamiento (e.g., García-Arroyo, & Osca, 2017) y autoeficacia laboral (Merino-Soto, Lima-Mendoza, Lozano-Huamán, Calderón de la Cruz, & Juárez-García, 2021); y en el desempeño laboral frente a la inseguridad laboral y afectividad negativa (Mäder, & Niessen, 2017). Sin embargo, es claro que la identificación de asociaciones no lineales monotónicas y no-monotónicas parece ser poco frecuente.

En contextos exploratorios donde se realiza la evaluación de dependencia estadística entre las variables, es especialmente útil detectar el tipo de relación funcional bivariada que existe (Clark, 2013; Reshef et al., 2011), sin necesariamente iniciar desde una hipótesis pre-definida sobre algún tipo de asociación, como ocurre habitualmente cuando se inicia con la búsqueda de relaciones lineales (Clark,

2013; de Siqueira Santos, Takahashi, Nakata, & Fujita, 2014; Dümcke, Mansmann, & Tresch, 2014; Fujita et al., 2009). Esto implica que la búsqueda de relaciones bivariadas debe iniciar con la detección de algún tipo de asociación existente, y que sirva como condición para realizar un segundo paso donde se prueben relaciones más específicas o más complejas (e.g análisis multivariados), ya sea desde un modelo lineal o no lineal. En este sentido, la incorporación de métodos modernos para la detección de un amplio rango de asociaciones bivariadas en ciencias de la conducta, son oportunidades aplicables también en la adaptación de medidas psicosociales del ambiente laboral.

Particularmente en el análisis de ítems de medidas de factores psicosociales del trabajo (FPST), existe evidencia de que la sintomatología de efectos psicológicos a nivel individual puede servir como marcador del efecto de factores psicosociales tales como el desbalance esfuerzo/recompensa, el limitado apoyo percibido, las altas demandas laborales, las demandas emocionales o el *mobbing* (o acoso laboral). Tales marcadores han sido, por ejemplo, los síntomas de ansiedad (Andrea, Bültmann, van Amelsvoort, & Kant, 2009; Duru, Ocaktan, Çelen, & Örsal, 2018; Magnavita, & Fileni, 2014), los síntomas de depresión (Duru et al., 2018; Magnavita, N., & Fileni, A. 2014; Berthelsen et al., 2015; Theorell et al., 2015; Żolnierczyk-Zreda, & Holas, 2018; Balducci, Avanzi, & Fraccaroli, 2014), y en general, el estado de salud mental (Duru et al., 2018; Harvey et al., 2017; Nieuwenhuijsen, Bruinvels, & Frings-Dresen, 2010; Stansfeld Candy, 2006). Por otro lado, la variabilidad individual derivada de constructos como la autoeficacia laboral, ha sido considerada como un factor moderador de la relación entre FPST y efectos positivos y negativos (Onyishi, Ugwu, Onyishi, & Okwueze, 2018); y el engagement (UWES), ha sido un indicador asociado con las demandas y recursos del trabajo (Merino-Soto et al., 2021; Reina-Tamayo, Bakker & Derks, 2017). En conjunto, estos constructos pueden ser buenos marcadores de la validez de los ítems cuando son considerados en la etapa de desarrollo y adaptación de una medida.

El análisis de ítems no estaría completo sin el examen de las características estructurales de los ítems, exactamente centrada en sus opciones de respuesta. Para este campo, existen heurísticas sobre las características óptimas para asegurar la calidad de las opciones de respuestas (e.g., Linacre, 1999, 2002), las que conjuntamente describen su funcionalidad y tienen influencia en el desempeño del puntaje y la validez de la estructura interna del instrumento (Smith, Wakely, de Kruif, & Swartz, 2003). Estas corresponden, por ejemplo, a la “atracción” de cada opción, la forma distribucional regular de respuestas, el número funcional de opciones, la monoticidad con relación al puntaje de la escala, y otras opciones, que pueden considerarse como condicionantes de la efectividad del sistema ordinal elegido (Engelhard, & Wind, 2012; Linacre, 2002; Smith et al., 2003). Dado que el funcionamiento de una escala debe converger con el óptimo funcionamiento de sus ítems, para hacer completo el estudio del contenido de los ítems, la validez de los ítems con criterios externos y el funcionamiento de sus opciones de respuesta, deben ser integradas en el análisis de los ítems.

En el presente estudio, los aspectos metodológicos anteriores (i.e., exploración de asociaciones no lineales, red nomológica y calidad del sistema ordinal) fueron aplicados a la adaptación al contexto peruano de una nueva medida factores y procesos laborales, la escala del *Instrumento para la Evaluación de Procesos Psicosociales del Trabajo* (PROPSIT; Juárez-García, & Flores-Jiménez, 2020; Paucar-Santivañez, 2020), creada para representar los principales factores del contexto laboral (e.g., demandas psicológicas, control, recompensas, etc.) factores extralaborales (e.g., estresores familiares, tráfico-traslado, etc), efectos psicológicos mediadores (e.g., burnout, engagement) y variables de salud mental relevantes en trabajadores mexicanos (síntomas somáticos, y alteraciones que incluyen depresión, ansiedad, estrés postraumático, entre otras). Desde el enfoque de la prevención primaria, las sub-escalas más importantes del PROPSIT es la de factores psicosociales laborales, en que se reconoce la existencia de factores negativos (factores de riesgo psicosocial) y la de factores positivos (recursos psicosociales favorables). El marco teórico de esta medida fue el modelo demandas-recursos (Bakker &

Demerouti, 2007; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001; Bakker & Demerouti, 2014), y está construido con una visión culturalmente relevante para su contexto mexicano de origen, pero es potencialmente generalizable a otros contextos latinoamericanos debido al análisis teórico de los factores con valor *etic* (i.e., factores consistentemente identificados en varios estudios internacionales; Broetje, Jenny, & Bauer, 2020), y que son consistentemente elegidos como atributos centrales para evaluar factores de riesgo psicosocial.

Es importante señalar que la clásica aproximación de determinación de propiedades psicométricas (e.g análisis factorial y estimación de confiabilidad) de medidas como el PROPSIT, aunque metodológicamente relacionada, puede ser independiente a la exploración psicométrica detallada del comportamiento de sus ítems en términos de la identificación de relaciones lineales o no lineales, la identificación de su red nomológica, y la calidad de estos en general. En un sentido estricto, los resultados pueden formar parte del estudio de la validez de contenido desde un ángulo cuantitativo, para evaluar las propiedades psicométricas a nivel de los ítems.

Dado lo anterior, el objetivo del presente estudio fue identificar las propiedades estructurales de los ítems de la escala de factores laborales del PROPSIT, así como sus asociaciones convergentes y divergentes con constructos de efectos psicosociales en trabajadores peruanos. Estas propiedades incluyeron, por un lado, la funcionalidad de las opciones de respuesta y su ordenamiento esperado, así como su función de distribución. Por otro lado, mediante un análisis asociativo de ítems con variables externas al modelo de medición del instrumento, se evaluó la validez de cada ítem con medidas de efectos psicosociales y procesos individuales teóricamente relevantes al modelo de medición del PROPSIT (red nomológica). Por ejemplo, la dimensión carga laboral puede esperarse que se asocie incrementalmente con insatisfacción laboral o estrés (Nieuwenhuijsen et al., 2010), y por lo tanto los ítems de esta medida deberían mantener asociaciones monotónicas positivas con medidas externas de insatisfacción laboral o estrés.

3.3.2. Método

Participantes

La muestra de participantes de obtuvo no probabilísticamente; los criterios de elegibilidad fueron: mayoría de edad, trabajador con contrato formal, y con actividad laboral actual (sin vacaciones, paro por enfermedad, etc). En la Tabla 1 aparecen las características de los trabajadores. La muestra convocada fue 188 trabajadores de varias ocupaciones y carreras, divididos en 69.7% trabajadores con estudios universitarios, 20.2% con estudios técnicos y 9.6% con estudios básicos. De acuerdo con la Clasificación Internacional Uniforme de Ocupaciones (CIUO-08; International Labour Organization, 2017), predominantemente (> 50%) los participantes mantenían ocupaciones técnicas de nivel medio (secretariado ejecutivo, técnico mecánico, cosmetología, panificación enfermería técnica, administración bancaria operarios, choferes) y profesionales (Tabla 16).

Instrumentos

Escala de Procesos Psicosociales del Trabajo (Juárez, 2018; Juárez-García, & Flores-Jiménez, 2020). Medida para evaluar las condiciones psicosociales laborales y los efectos sobre el bienestar del trabajador. Contiene cuatro secciones generales: Factores Psicosociales del Trabajo (FPST), Efectos psicosociales del trabajo, Consecuencias de Salud-Enfermedad y Otros Factores Psicosociales extra-laborales. El presente estudio validó la sección de FPST, consistente de dos grandes áreas: *Recursos Psicosociales Favorables* (dimensiones de recompensas y desarrollo laboral, control laboral, recursos para hacer el trabajo, clima y apoyo social laboral y congruencia de valores), y *Factores de Riesgo Psicosociales*, que consiste en varias dimensiones como Demandas Laborales (subdimensiones: carga y ritmo de trabajo, alta responsabilidad y peligrosidad, jornadas, turnos u horarios, cognitivas o atencionales, emocionales por trato con personas, esfuerzo físico y ambiente físico), Acoso Laboral psicológico y Liderazgo Estresante.

Tabla 16

Descripción de la muestra efectiva del estudio (n = 188)

	N	%
Sexo		
Hombre	96	51,1
Mujer	92	48,9
Estado civil		
Casado(a)	37	19,7
Conviviente	24	12,8
Divorciado(a)	3	1,6
Soltero(a)	124	124
Contrato		
Estable/Definitivo	95	50,5
Eventual/Temporal	93	49,5
Jerarquía		
Director, gerente, supervisor o jefe de área	27	14,4
Administrativo, sin personal a cargo (trabajo no manual)	101	53,7
Ventas sin personal a cargo (trabajo no manual)	9	4,8
Empleado operativo (trabajo manual)	49	26,1
Otro	2	1,1
Convivencia		
Con amigos	3	1,6
Con pareja o familiares	171	91,0
Otro	2	1,1
Solo	11	5,9
Solo con mascota	1	,5
Profesión		
Ciencias de la salud	6	3,2
Ciencias básicas	1	,5
Ingeniería	20	10,6
Ciencias económicas y de la gestión	65	34,6
Humanidades, ciencias jurídicas y sociales	39	20,7
No pertenece/ no aplica	56	29,8
Sin dato	1	,5
Clasificación CIUO-08		
Directores y gerentes	3	1,6
Profesionales científicos e intelectuales	58	30,9
Técnicos y profesionales de nivel medio	57	30,3
Personal de apoyo administrativo	30	16,0
Servicios, vendedores de comercios y mercados	8	4,3
Oficiales, operarios, artesanos y artes mecánicas y otros oficios	26	13,8
Operadores de instalaciones, máquinas y ensambladores	4	2,1
Ocupaciones elementales	1	,5

Escala de autoeficacia ocupacional (OSES; Rigotti, et al., 2008). Instrumento de seis ítems que evalúa la autoeficacia en el contexto laboral global, mediante la percepción de suficiencia y confianza en la habilidad para eventos laborales. Su alta magnitud de confiabilidad y relación con criterios laborales y otros constructos fue corroborada en contexto peruano (Merino-Soto, Lima-Mendoza, Lozano-Huamán, Calderón de la Cruz, & Juárez-García, 2021). En el presente estudio la confiabilidad fue satisfactoria, $\omega = .95$ (IC 95% = .92, .96).

Escala de Engagement de Utrech (UWES-3; Schaufeli, Shimazu, Hakanen, Salanova, & De Witte, 2019). Evalúa el entusiasmo laboral, con tres ítems sobre vigor, dedicación y absorción elegidos con base racional y empírica por los autores y derivados del UWES-9. El escalamiento de respuesta es de siete puntos, entre 0 (*nunca*) y 6 (*siempre*). En dos estudios peruanos, los resultados de validez confirmaron su validez de constructo respecto a la estructura interna, confiabilidad y relación con otras variables (Calderón-de la Cruz et al., en preparación; Merino-Soto et al., en preparación). En el presente estudio, la estructura interna fue adecuada, $\omega = .94$ (IC 95% = .91, .96).

Ítem Único de Estrés (IUE; Elo, Leppänen, & Jahkola, 2003). Medida de un solo ítem creada para evaluar los síntomas de estrés en un marco de referencia global, escalado ordinalmente en cinco opciones (desde nada hasta mucho). En general, es medida eficiente para aproximarse a la experiencia del estrés, y en particular ha sido utilizado en estudios sobre estrés laboral (Arapovic-Johansson, Wåhlin, Kwak, Björklund, & Jensen, 2017; Elo et al., 2003; Houdmont et al., 2019; Omholt, Tveito, & Ihlebæk, 2017; Salminen et al., 2014)

Patient Health Questionnaire-4 (PHQ-4; Kroenke, Spitzer, Williams, & Löwe, 2009). Medida breve de despistaje de síntomas emocionales y cognitivos de depresión (dos ítems) y ansiedad generalizada (ítems), internacionalmente aceptada como una medida de despistaje total de eficiente distrés psicológico. Está escalada con cinco puntos (desde *para nada* hasta *casi todos los días*) en un marco de las últimas dos semanas. En estudios con participantes peruanos (Merino-Soto, Dominguez-Lara, & Fernández-Arata, 2017; Merino-Soto, Angulo-Ramos, & López-Fernández, 2019), sus correlaciones mantienen coherencia teórica con otros constructos. En los presentes datos, el ajuste a dos dimensiones correlacionadas (i.e., ansiedad y depresión) fue satisfactorio, $WLSMV-\chi^2 = .067$ (gl = 1, $p = .79$) CFI = 1.00, SRMR = .00, WRMR = .061, pero la correlación inter-factorial fue muy alta ($r = .94$, IC 95% = .88, .99); por lo tanto, se utilizó un puntaje único ($WLSMV-\chi^2 = 3.01$, gl = 2, $p = .22$, CFI = 1.00, SRMR = .02, WRMR = .409; cargas factoriales: .89, .90, .95 y .93), interpretado como

distrés psicológico (Kroenke et al., 2009); la consistencia interna del puntaje fue $\alpha = .90$ (IC 95% = .85, .93).

Procedimiento.

Recolección de datos. La convocatoria a los participantes se hizo mediante las redes sociales del investigador principal, y dos colaboradoras peruanas, egresadas de la carrera de psicología; las redes sociales usadas fueron WhatsApp y Facebook. La elegibilidad de la selección de los contactos se estableció en la edad (18 años mínimo) y en la actividad laboral (estar trabajando en el momento de la recolección de datos). El primer contacto con los participantes fue para explicar brevemente por escrito, en qué consistió el estudio y el contenido de la encuesta; luego, los participantes recibieron el vínculo de la encuesta por el mismo medio de comunicación. Los participantes fueron también invitados a reenviar el vínculo a sus propios contactos. El material fue ordenado en una secuencia única para todos: formulario de consentimiento de participación (con información sobre el objetivo del estudio, la voluntariedad de la participación, la anonimidad de respuesta, la libertad para dejar de responder en cualquier momento, el tratamiento confidencial de los datos, y la ausencia de rastreabilidad del participante). La elegibilidad de los participantes luego de recolectar los datos se orientó a) la pertenencia contractual a una institución laboral peruana en el último empleo, y b) el tiempo de trabajo de mínimo de seis meses.

Análisis. Se obtuvieron medidas descriptivas univariadas, estructurales (uso de categorías de respuestas, similaridad de respuesta, distribución del escalamiento) y asociativas (relación con sexo, edad, autoeficacia ocupacional, engagement y distrés psicológico). Las características estructurales tuvieron como marco de referencia los aspectos necesarios para establecer la calidad estructural de los ítems (Engelhard & Wind, 2012; Linacre, 2002; Smith et al., 2003).

Propiedades descriptivas. Se obtuvieron medidas de tendencia o intensidad de respuesta, asimetría, curtosis y normalidad estadística univariadas, examinada con la prueba *Anderson-Darling* en

el nivel nominal .15 (Keselman, Othman, & Wilcox, 2013); esta prueba puede ser considerada eficiente para variables ordinales (Keselman et al., 2013). Se utilizó el programa *MVN* (Korkmaz S, Goksuluk D, Zararsiz, 2014).

Eficiencia del uso de las categorías de respuesta. En primer lugar, se registró frecuencia absoluta de respuestas en cada categoría, y fueron señaladas como problemáticas o ineficientes las categorías con menos de 10 observaciones (Linacre, 2002). Este número es el tamaño mínimo recomendado para garantizar la precisión y estabilidad de las estimaciones psicométricas en cada opción de respuesta (Engelhard, & Wind, 2012; Smith et al., 2003). En segundo lugar, se identificó el número de opciones de respuestas efectivas (*actual equivalent number of options*: AENO; Sato, & Morimoto, 1976). También llamado *número de alternativas hipótéticas equivalentes* (Samejima, 1980), es un indicador cuantitativo del número de opciones funcionales que efectivamente ocurren en la muestra. Está formulado en la conceptualización de la entropía, mediante:

$$\text{AENO} = 2^{-\sum_{i=1}^k p_i \log_2(p_i)}$$

En la fórmula, k es el número de opciones, p es la proporción de respuesta en la opción k , \log es logaritmo con base 2. AENO representa la cantidad de información máxima definida como el número de distintas opciones de respuestas. Las opciones cuyas frecuencias fueron cero fueron tratadas como ceros incidentales (Linacre, 2002), fueron reemplazadas por .05.

Distribución del escalamiento. Los ítems fueron tratados como variables categóricas ordinales para examinar sus propiedades distribucionales. Primero, se aplicó el enfoque *AJUS* (Galton, 1969) para identificar el patrón de respuestas mediante cuatro categorías clasificatorias: *A* (i.e., unimodal), *J* (i.e., unimodal con elevación en alguna de las colas), *U* (i.e., bimodal con elevación en ambas colas), y *S* (i.e., multimodal).

Segundo, el grado de similaridad de respuesta dentro del mismo ítem fue estimado por una medida de consenso para variables ordinales, coeficiente *TW* (varía entre 0 y 1.0; Tastle, & Wierman, 2007); valores cercanos a 1.0 indican mayor consenso o similaridad de respuesta. Para estos análisis se utilizó el programa *agrmt* (Ruedin, 2020).

Asociación con variables sociodemográficas. Para detectar posible covariación respecto al sexo, se utilizó la *correlación biserial por rangos* (r_{br} ; Glass, 1966), implementada con el programa *wilcoxonRG* (Mangiafico, 2020), con intervalo de confianza de 95%. Con respecto a la edad, para detectar alguna forma de dependencia monótonica, se utilizó el coeficiente de correlación de Spearman

Asociación para la validez de constructo. Para identificar las relaciones convergentes y divergentes de los ítems, se utilizaron los puntajes del OSES (i.e., autoeficacia ocupacional), entusiasmo (i.e., UWES), las medidas de estrés (i.e., ítem único de estrés), y de síntomas de distrés psicológico o salud mental (i.e., PHQ-4) como criterios. Para detectar dependencias en amplio rango de covariabilidad (e.g., monótonicas lineales, monótonicas no lineales, y no monótonicas) se utilizó el *coeficiente de máxima información* (MIC; Reshef et al., 2018), que proviene del enfoque *Maximal Information-Based Nonparametric Exploration* (MINE, Reshef et al., 2011); comparado con otras medidas asociacionales (i.e., correlación distancia) esta medida no paramétrica muestra adecuado poder para detectar un amplio rango de estructuras de dependencia (de Siqueira Santos, Takahashi, Nakata, & Fujita, 2014; Dümcke, Mansmann, & Tresch, 2014). MIC provee un valor entre 0.0 (completa independencia estadística) hasta 1.0 (dependencia completa) indicando la fuerza de la relación en para algún tipo de relación función bivariada (no necesariamente lineal). Debido que puede ser comparado con el porcentaje de covarianza estandarizada (r^2) y que medidas de información mutua tienden a ser ineficientes para detectar relaciones lineales (Clark, 2013; Makonnen, 2019), se utilizó juntamente con la correlación lineal para detectar la linealidad de las relaciones. Se usó el programa *testforDEP* (Miecznikowski, Hsu, Chen, & Vexler, 2017).

3.3.3. Resultados

Propiedades descriptivas de los ítems

Sección Factores de Riesgo Psicosocial, FRPS (Tabla 17). En general, la intensidad de respuesta (respuesta media) varió consistentemente el rango de las opciones “*Nunca*” y “*Regularmente (algunas veces al mes)*”, excepto en las áreas *Cognitivas / atencionales*, *Carga y ritmo de trabajo* y *Retroalimentación negativa*, donde la tendencia fue mayor. Por otro lado, la diferencia de respuesta media dentro de cada área de contenido fue notoria en *Carga y ritmo de trabajo* y *Jornada, turno y horario*, sugiriendo contenidos conceptualmente heterogéneos. Excepto en estos contenidos, en general la tendencia de respuesta fue similar en las propias áreas de contenido de factores de riesgo psicosocial. La variabilidad fue consistente dentro de cada contenido, y en general varió moderadamente ($DE_{\max} / DE_{\min} = 2.38$).

La tendencia de la asimetría y curtosis fue consistente en cada área de contenido, excepto en algunas donde hubo observable variabilidad en la dirección (e.g., *jornada, turno y horario*). La normalidad estadística (estadístico AD) no fue aceptada en ningún ítem, y fue de similar magnitud en cada área de contenido; la excepción fue *jornada, turno y horario*, en donde el grado de no normalidad varió considerablemente. Por otro lado, el grado de acuerdo de respuesta fue moderadamente alto ($T_w > .50$) indicando la predominancia de distribuciones unimodales y regulares.

Sección Recursos Favorables, RPSF (Tabla 18). La tendencia de respuesta predominante estuvo concentrada en las opciones *frecuentemente* y *muy frecuentemente*, sugiriendo que la densidad de respuesta se orienta hacia la derecha del escalamiento, algo congruente con las distribuciones asimétricas negativas detectadas. Esta forma de asimetría ocurrió en todos los ítems (excepto en los

ítems que se recodifican), y la curtosis indicó ser comparativamente más variable entre los ítems; en conjunto, la normalidad estadística no fue cumplida por alguno de los ítems.

Uso de las categorías de respuesta

Sección Factores de Riesgo Psicosocial, FRPS (Tabla 17). Se observó que el número de categorías en los 22 ítems de FRPS (factores de riesgo psicosocial) con menos de 10 opciones ($O < 10$) y AENO, correlacionaron linealmente en -0.928 , sugiriendo alta convergencia entre ambas métricas. Excepto cuatro ítems (18.1%; ítems 4, 9, 12 y 22) en FRPS con respuestas eficientes entre 1 y 7 opciones, el resto produjo entre uno y cinco opciones con menos de 10 respuestas. La alta tasa de opciones no utilizadas ocurrió en los ítems referidos a la interacción social entre trabajadores (i.e., *acoso laboral y liderazgo estresante*) y el horario de actividad (*jornadas, turno y horario*).

Sección Recursos Psicosociales Favorables, RPSF (Tabla 18). Las opciones eficientes de respuesta AENO variaron entre 3.8 (ítem 27) y 5.99 (ítem 24), y predominantemente con 5 opciones eficientes ($M_{AENO} = 4.99$). Cerca de la mitad de los ítems (11, 57.8%) produjeron solo una opción de respuesta ineficiente, mientras que el resto produjo entre dos y cuatro frecuencias debajo de 10 respuestas. Comparado con la sección de factores de riesgo psicosocial, esta sección contiene menos opciones ineficientes. La concordancia entre el $O < 10$ y AENO en los 19 de ítems RPSF (recursos favorables) fue $r = -0.883$ ($p < .05$), que similar a lo encontrado en la sección de anterior (Factores de Riesgo Psicosocial), esta asociación puede considerarse fuerte.

Distribución del escalamiento

Sección Factores de Riesgo Psicosocial, FRPS. Los resultados de presentan en la Tabla 17 . El sistema AJUS identificó la siguiente proporción de formas distribucionales: 6 (27.2%), 0, 4 (18.1%), 10 (45.4) y 2 (9%) para A, J, U, S y L, respectivamente. Los ítems identificados con la forma distribucional S (multimodal) fue predominante, indicando posibles procesos diferentes de respuesta en

estos ítems; sin embargo, se detectó discrepancia entre este tipo de distribución y su alta concentración de respuesta (medida por T_w). Para verificar esta aparente multimodalidad, se vio la correspondencia entre la observación de los gráficos de densidad distribucional (no reportados aquí por limitaciones de espacio) y los coeficientes T_w , y se halló una clara regularidad de la distribución de respuesta de los ítems.

Sección Recursos Psicosociales Favorables, RPSF. Los resultados hallados se muestran en la Tabla 18. Las formas distribucionales fueron identificadas con la siguiente frecuencia: A = 6 (31.5%); J = 1 (5.2%); U = 0; S = 11 (57.8%), y L = 1 (5.2%), indicando que cerca de la mitad tuvo una aparente distribución irregular. Como en la sección de factores psicosociales anterior, la predominancia de las distribuciones multimodales (“S”) discrepó de la concentración de respuesta medida por T_w : en estos ítems T_w osciló entre .55 y .72 indicando moderados o elevada concentración de respuesta. Sin embargo, la observación de gráficos de densidad en estos ítems con aparente irregularidad distribucional mostró que las distribuciones pueden reconocerse como unimodales.

Tabla 17

Resultados descriptivos para los ítems de Factores de Riesgo Psicosocial del PROPSIT

FACTORES DE RIESGO PSICOSOCIAL (RP)	M	DE	As	Cu	AD	AJUS	Tw	O < 10	AENO
RP: Demandas laborales									
Carga y ritmo de trabajo									
1. Carga de trabajo	4.67	1.53	-.57	-.85	9.96	A	.62	1, 7	5.29
2. Trabajar rápido	4.74	1.56	-.64	-.57	8.24	S	.62	1	5.8
3. Interrupciones para tareas a tiempo	2.83	1.44	.94	.38	9.05	A	.66	7	5.17
Alta responsabilidad y peligrosidad									
4. Tomar acciones importantes	2.54	1.72	1.00	-.09	13.06	L	.56	0	5.20
5. Peligros	2.07	1.32	1.53	2.22	15.60	U	.71	5, 6, 7	4.07
Jornada, turno y horario									
6. Horas extras, horarios prolongados	3.30	1.49	.28	-.88	5.36	S	.64	7	5.71
7. Rotación/cambio de turno	1.72	1.31	1.93	2.94	33.09	S	.71	4, 6, 7	2.99
8. Debo laborar jornadas en turno nocturno	1.58	1.24	2.45	5.81	4.22	U	.73	4, 5, 6, 7	2.55
Cognitivas / atencionales									
9. Atención prolongada en tareas	4.53	1.74	-.55	-.92	8.76	S	.55	0	5.95
10. Atender dos o más tareas	4.64	1.66	-.59	-.79	9.04	A	.58	1	5.75
11. Esfuerzo mental	4.60	1.65	-.37	-.96	6.00	A	.58	1	6.09
Emocionales									
12. Emociones negativas de otras personas	3.51	1.58	.62	-.39	6.39	A	.62	0	5.88
13. Trato con personas desgastante	2.51	1.25	1.14	1.47	1.27	S	.72	6, 7	4.45
14. Mostrar emociones diferentes a las que siento	2.64	1.55	1.15	.65	11.73	A	.63	6, 7	5.09
Esfuerzo físico									
15. Mucho esfuerzo físico	2.53	1.66	1.01	-.10	13.39	S	.59	7	5.04
16. Posiciones incómodas	2.28	1.53	1.18	.19	17.30	S	.54	7	4.35
Ambiente físico									
17. Condiciones ambientales adversas	2.17	1.34	1.21	.83	13.79	L	.70	6, 7	4.34
RP: acoso laboral									
Por superiores									
18. Maltrato psicológico (superiores)	1.26	0.76	4.39	22.30	47.61	U	.80	3, 4, 5, 6, 7	1.80
Por compañeros									
19. Maltrato psicológico (compañeros)	1.21	0.73	5.14	29.29	53.60	U	.81	3, 4, 5, 6, 7	1.59
RP: Liderazgo estresante									
Supervisión controlante									
20. Demasiado control	1.84	1.02	2.08	5.95	17.80	S	.80	4, 5, 6, 7	3.18
21. Reglas y normas exageradas	1.96	1.14	1.73	3.22	16.77	S	.72	4, 5, 6, 7	3.52
Retroalimentación negativa (supervisor)									
22. Retroalimentación adecuada de supervisores (R)	4.69	1.69	.75	-.48	9.59	S	.58	0	5.71

Nota. RP: Riesgo Psicosocio. As y Cu: coeficientes de asimetría y curtosis. AD: prueba de normalidad Anderson-Darling. AJUS: clasificación de distribuciones. Tw: coeficiente de concentración. AENO: actual equivalent number of options.

Tabla 18

Resultados descriptivos para los ítems de Recursos Psicosociales Favorables del PROPSIT

	M	DE	As	Cu	AD	AJUS	Tw	O < 10	AENO
RECURSOS PSICOSOCIALES FAVORABLES (RF)									
RF: Recompensas y desarrollo laboral									
Justicia organizacional									
23. Trabajo justo y equitativa.	4.85	1.58	-.72	-.41	9.19	S	.61	7	5.48
Salario motivante									
24. Salario motivante	4.74	1.76	-.63	-.77	8.91	S	.55	0	5.99
Reconocimiento por el trabajo									
25. Trabajo apreciado y reconocido	4.92	1.46	-.75	-.26	9.08	S	.66	1	5.24
Posibilidades de desarrollo laboral/profesional (estatus)									
26. No oportunidades de crecer profesionalmente (R)	2.27	1.33	1.05	.58	11.25	L	.69	5, 6, 7	4.50
Seguridad en la conservación del empleo									
27. Pienso que puedo perder este empleo próximamente (R)	2.04	1.26	1.66	2.70	16.12	S	.73	4, 5, 6, 7	3.88
Tarea gratificante y con trascendencia									
28. Trabajo agradable y gratificante	5.20	1.43	-.89	.11	9.67	A	.66	1	5.09
29. Tareas benefician a personas y sociedad	5.28	1.40	-.90	.20	9.18	A	.67	1, 2	5.07
RF: Control laboral y contenido de la tarea									
Influencia – autonomía en el trabajo									
30. Libre en decidir el trabajo	5.18	1.48	-.94	.13	10.23	A	.65	1	5.16
Utilización de habilidades en el trabajo									
31. Utilizo habilidades	5.68	1.38	-1.17	.64	13.32	S	.68	1, 2	4.49
32. Desarrollo de capacidades	5.67	1.42	-1.08	.32	12.92	J	.66	1, 2	4.95
Trabajo variado (no monótono)									
33. Actividades muy variadas	5.41	1.49	-.85	-.32	11.40	S	.63	1, 2	4.60
Funciones y roles claros									
34. Funciones y tareas muy claras	5.62	1.39	-1.16	.75	12.59	S	.67	1, 2	4.64
RF: Recursos para realizar el trabajo									
Recursos materiales, equipos, para el trabajo									
35. Materiales necesarios y adecuados	5.58	1.49	-1.28	.95	14.34	S	.65	1	5.49
Capacitación, adiestramiento									
36. Capacitaciones necesarias	5.01	1.64	-.91	-.10	10.46	S	.60	1	4.99
RF: Clima y apoyo social laboral									
Apoyo de compañeros									
37. Compañeros apoyan	5.10	1.36	-.80	.32	8.23	A	.70	1	5.11
Apoyo de jefes									
38. Superiores apoyan	5.05	1.34	-.57	-.08	6.68	A	.70	1	4.93
Clima de unión									
39. Clima de unión/colaboración	5.22	1.34	-.79	.18	8.19	S	.69	1	4.87
RF: congruencia de valores									
Individuo - organización									
40. Coincidencias mis valores y los de la organización	5.09	1.36	-.80	.03	9.11	S	.62	1	5.17
Individuo – compañeros – superiores									
41. Coincidencia mis valores – compañeros	4.96	1.41	-.87	.20	8.96	A	.68	1	5.17

Nota. RF: Recurso Favorable. As y Cu: coeficientes de asimetría y curtosis. AD: prueba de normalidad Anderson-Darling. AJUS: clasificación de distribuciones. Tw: coeficiente de concentración. AENO: actual equivalent number of options.

Asociación con variables externas

Sección de Factores de Riesgo Psicosocial (FRPS). El sexo y la edad mantuvieron asociaciones alrededor de cero o debajo de .18, y todas estadísticamente no significativas (Tabla 19). Con respecto a las relaciones esperadas con otros constructos, excepto los ítems de demandas cognitivas-atencionales (atención prolongada a tareas, a dos o más tareas y elevado esfuerzo mental), y demandas de carga y ritmo de trabajo (monto y rapidez), el resto de los contenidos mostraron consistencia teórica, es decir, asociación negativa con autoeficacia (OSES; $r_{mediana} = -.22$), engagement (UWES; $r_{mediana} = -.22$), y positivas con alteraciones de salud mental (PHQ-4; $r_{mediana} = .16$). Con el ítem único de estrés, la dirección fue teórica y consistentemente positiva ($r_{mediana} = .10$). En el contenido carga y ritmo de trabajo, el ítem 3 (interrupciones para terminar las tareas a tiempo) fue inconsistente respecto a los otros dos ítems de su área de contenido; similarmente, el ítem 6 (horas extras) de *jornada, turno y horario* mostró inconsistencia respecto a los otros dos ítems de su área de contenido.

Sección Recursos Psicosociales Favorables (RPSF). El sexo y la edad mantuvieron asociaciones alrededor de cero y estadísticamente no significativas (Tabla 20). La magnitud mediana de las asociaciones monotónicas se mantuvo en el siguiente orden: UWES ($r_{mediana} = .56$), OSES ($r_{mediana} = .43$), PHQ-4 ($r_{mediana} = -.35$) y estrés ($r_{mediana} = -.03$). En general, a) la dirección de la asociación fue teóricamente consistente de todos los ítems con cada variable externa, y b) se halló que las asociaciones con todas las variables externas no variaron sustancialmente: en UWES, OSES y PHQ el tamaño predominante alto puede considerarse grande tomando en cuenta la naturaleza categórica de las variables. Con la medida de estrés, la asociación monotónica fue trivial.

Tabla 19

Resultados asociativos para los ítems de Factores de Riesgo Psicosocial

FACTORES DE RIESGO PSICOSOCIAL (RP)	IUE (estrés)		OSES (autoeficacia)		UWES (engagement)		PHQ-4 (distrés)		Sexo	Edad
	MIC	<i>rho</i>	MIC	<i>rho</i>	MIC	<i>rho</i>	MIC	<i>rho</i>		
RP: Demandas laborales										
<i>Carga y ritmo de trabajo</i>										
1. Carga de trabajo	.083	<i>.140</i>	.136	<u>.222</u>	.159	<u>.343</u>	.204	<u>-.289</u>	.046	.069
2. Trabajar rápido	.101	<i>.175</i>	.142	<i>.144</i>	.119	<u>.261</u>	.173	<u>-.355</u>	.076	.006
3. Interrupciones para tareas a tiempo	.061	.047	.190	<u>-.305</u>	.155	<u>-.407</u>	.245	<u>.486</u>	-.104	-.006
<i>Alta responsabilidad y peligrosidad</i>										
4. Tomar decisiones (vida o seguridad)	.045	.00	.104	<i>-.176</i>	.103	<u>-.225</u>	.056	<u>.23</u>	.112	.037
5. Peligros	.038	<i>-.027</i>	.152	<u>-.292</u>	.138	<u>-.303</u>	.082	<i>.175</i>	<i>.171</i>	.015
<i>Jornada, turno y horario</i>										
6. Horas extras, horarios prolongados	.05	<i>.097</i>	.116	.007	.107	<i>.140</i>	.071	<i>-.178</i>	-.029	-.072
7. Rotación/cambio de turno	.043	.067	.146	<u>-.303</u>	.164	<u>-.290</u>	.084	<i>.142</i>	.00	-.002
8. Laborar jornadas en turno nocturno	.052	.024	.160	<u>-.343</u>	.156	<u>-.353</u>	.107	<u>.283</u>	.074	.058
<i>Cognitivas / atencionales</i>										
9. Atención prolongada en tareas	.086	.096	.169	<u>.285</u>	.152	<u>.319</u>	.206	<u>-.420</u>	.085	.110
10. Atender dos o más tareas	.082	<i>.140</i>	.134	<u>.243</u>	.146	<u>.290</u>	.191	<u>-.361</u>	.076	.011
11. Esfuerzo mental	.041	.100	.228	<u>.416</u>	.181	<u>.439</u>	.122	<u>-.295</u>	.019	.089
<i>Emocionales</i>										
12. Emociones negativas otras personas	.061	.046	.129	<i>-.042</i>	.112	<i>-.117</i>	.099	<u>.242</u>	<i>-.176</i>	.028
13. Trato con personas desgastante	.039	<i>.175</i>	.139	<u>-.254</u>	.110	<u>-.249</u>	.09	<u>.214</u>	-.110	.016
14. Mostrar emociones diferentes	.064	<i>.176</i>	.150	<u>-.324</u>	.150	<u>-.353</u>	.232	<u>.432</u>	-.145	-.006
<i>Esfuerzo físico</i>										
15. Mucho esfuerzo físico	.052	<i>.164</i>	.143	<u>-.188</u>	.112	<u>-.217</u>	.064	.135	.103	.018
16. Posiciones incómodas	.061	.117	.133	<u>-.212</u>	.119	<u>-.212</u>	.059	.129	.082	-.047
<i>Ambiente físico</i>										
17. Condiciones ambientales deficientes	.070	.120	.187	<u>-.229</u>	.107	<u>-.192</u>	.129	<u>.204</u>	<i>.168</i>	.005
RP: acoso laboral										
<i>Por superiores</i>										
18. Superiores, maltrato psic. Por compañeros	.028	<i>.137</i>	.184	<u>-.347</u>	.117	<u>-.307</u>	.107	<u>.246</u>	.005	.039
19. Compañeros me maltrato psic.	.024	.036	.114	<u>-.256</u>	.117	<u>-.272</u>	.089	<u>.216</u>	.005	.118
RP: Liderazgo estresante										
<i>Supervisión controlante</i>										
20. Demasiado control	.050	<i>.144</i>	.143	<u>-.274</u>	.106	<u>-.253</u>	.129	<i>.151</i>	-.012	.121
21. Reglas y normas exageradas	.044	.115	.124	<u>-.233</u>	.151	<u>-.288</u>	.112	<i>.18</i>	<i>.154</i>	.056
<i>Retroalimentación negativa del supervisor</i>										
22. Retroalimentación adecuada (R)	.078	.080	.194	<u>.345</u>	.290	<u>.549</u>	.234	<u>-.397</u>	-.007	-.202

Nota. IUE: ítem único de estrés. OSES: autoeficacia ocupacional. UWES: engagement. PHQ-4: distrés psicológico. MIC: coeficiente de máxima información. *r*: correlación monótonica Spearman. En cursiva: $p < .05$. Subrayado: $p < .002$ (corrección Bonferroni en $\alpha = .05$)

Tabla 20

Resultados asociativos para los ítems de Recursos Psicosociales Favorables

RECURSOS PSICOSOCIALES FAVORABLES (RF)	IUE (estrés)		OSES (autoeficacia)		UWES (engagement)		PHQ-4 (distrés)		Sexo	Edad
	MIC	rho	MIC	rho	MIC	rho	MIC	rho		
RF: Recompensas y desarrollo laboral										
<i>Justicia organizacional</i>										
23. Trabajo justo y equitativa.	.057	.00	.269	<u>.443</u>	.267	<u>.561</u>	.212	<u>-.394</u>	.090	-.060
<i>Salario motivante</i>										
24. Salario motivante	.080	-.030	.222	<u>.437</u>	.298	<u>.649</u>	.308	<u>-.579</u>	.110	-.040
<i>Reconocimiento por el trabajo</i>										
25. Trabajo apreciado y reconocido	.060	.017	.264	<u>.447</u>	.291	<u>.612</u>	.201	<u>-.358</u>	.06	-.1
<i>Posibilidades de desarrollo laboral</i>										
26. No oportunidades de crecer profesionalmente (R)	.068	.00	.161	<u>-.333</u>	.204	<u>-.447</u>	.158	<u>.419</u>	-.070	.140
<i>Seguridad en la conservación del empleo</i>										
27. Pienso que puedo perder este empleo próximamente (R)	.032	-.038	.171	<u>-.326</u>	.171	<u>-.408</u>	.138	<u>.277</u>	.060	-.060
<i>Tarea gratificante y con trascendencia</i>										
28. Trabajo agradable y gratificante	.046	-.048	.239	<u>.457</u>	.394	<u>.674</u>	.21	<u>-.378</u>	.020	.00
29. Tareas benefician a personas y sociedad	.040	.041	.141	<u>.353</u>	.247	<u>.433</u>	.086	<u>-.186</u>	-.070	-.030
RF: Control laboral y contenido de la tarea										
<i>Influencia – autonomía en el trabajo</i>										
30. Libre en decidir el trabajo	.046	-.057	.167	<u>.411</u>	.298	<u>.513</u>	.136	<u>-.26</u>	.00	.060
<i>Utilización de habilidades en el trabajo</i>										
31. Utilizo habilidades	.027	.00	.184	<u>.477</u>	.389	<u>.634</u>	.189	<u>-.371</u>	0	-.040
32. Desarrollo de capacidades	.028	.013	.231	<u>.511</u>	.385	<u>.684</u>	.183	<u>-.376</u>	.010	-.020
<i>Trabajo variado (no monótono)</i>										
33. Actividades muy variadas	.020	.027	.219	<u>.513</u>	.321	<u>.652</u>	.165	<u>-.395</u>	.010	.020
<i>Funciones y roles claros</i>										
34. Funciones y tareas muy claras	.047	-.010	.185	<u>.458</u>	.340	<u>.632</u>	.209	<u>-.393</u>	-.050	.030
RF: Recursos para realizar el trabajo										
<i>Recursos materiales, etc., para trabajar</i>										
35. Materiales necesarios y adecuados	.082	-.110	.169	<u>.435</u>	.345	<u>.582</u>	.241	<u>-.421</u>	.080	.00
<i>Capacitación, adiestramiento</i>										
36. Capacitaciones necesarias	.086	-.067	.253	<u>.454</u>	.298	<u>.591</u>	.239	<u>-.407</u>	.040	-.070
RF: Clima y apoyo social laboral										
<i>Apoyo de compañeros</i>										
37. Compañeros apoyan	.066	-.090	.195	<u>.310</u>	.293	<u>.453</u>	.107	<u>-.177</u>	.050	-.120
<i>Apoyo de jefes</i>										
38. Superiores apoyan	.057	-.052	.188	<u>.354</u>	.274	<u>.478</u>	.132	<u>-.159</u>	.03	-.100
<i>Clima de unión</i>										
39. Clima de unión/colaboración	.049	-.073	.227	<u>.365</u>	.303	<u>.52</u>	.141	<u>-.259</u>	.130	-.030
RF: congruencia de valores										
<i>Congruencia individuo – organización</i>										
40. Valores propios y organización individuo – compañeros – superiores	.047	-.051	.22	<u>.385</u>	.344	<u>.544</u>	.161	<u>-.304</u>	.090	-.070
41. Coincidencia mis valores – compañeros	.053	.036	.17	<u>.348</u>	.253	<u>.497</u>	.17	<u>-.289</u>	.03	-.07

Nota. IUE: ítem único de estrés. OSES: autoeficacia ocupacional. UWES: engagement. PHQ-4: distrés psicológico. MIC: coeficiente de máxima información. r: correlación monotónica Spearman. En cursiva: $p < .05$. Subrayado: $p < .002$ (corrección Bonferroni en $\alpha = .05$)

3.3.4. Discusión

En el marco de una de las etapas de evaluación de la validez de contenido a nivel de los ítems, en el presente estudio se presentaron por primera vez evidencias del PROPSIT (Juárez-García, & Flores-Jiménez, 2020; Paucar-Santivañez, 2020) para su posible aplicación al contexto peruano, y dado el enfoque ético de sus constructos, con potencial aplicación a otros contextos. Esta medida fue creada originalmente para el contexto mexicano y representa un avance en la integración de los modelos de factores psicosociales del trabajo (FPT) con más evidencias internacionales (desequilibrio esfuerzo-recompensa, Siegrist, 1996; demanda-control, Karasek, & Theorell, 1990; demandas-recursos, Bakker, & Demerouti, 2014), y la adecuación cultural provenientes de los FPT en México (Juárez-García, & Flores-Jiménez, 2020) pero con potencial generalización intercultural. En general, el presente estudio expone evidencias en un conjunto de propiedades que usualmente no son reportadas en la validación de una medida aplicada la medición de los factores psicosociales del trabajo; estas propiedades se enfocaron en los ítems respecto a una parte de sus características estructurales y la asociación con constructos externos.

Sobre la tendencia de respuesta en los ítems, se observó un claro patrón donde los factores de riesgo psicosocial fueron menos intensas comparados con los factores de recursos favorables, indicando que la predominancia de un conjunto de factores sobre otros, y la capacidad de los ítems para diferenciarlos. Esta diferenciación global también sugiere que aún con la variabilidad de puestos y contextos de trabajo muestreados, la tendencia identifica ambientes donde los recursos favorables son percibidos como más frecuentes que los factores de riesgo psicosocial.

En relación con el escalamiento, las respuestas no se distribuyeron normalmente y esta distorsión provino del tercer y cuarto momento de la distribución (i.e., asimetría y curtosis). El grado de distorsión indicado por estos momentos no parece ser fuerte, y está dentro de los límites no excesivos de curtosis y asimetría (Brown, 2015). Esta no normalidad parece ser la regla y no excepción

característica de las mediciones en ciencias sociales (Bono, Blanca, Arnau, & Gómez-Benito, 2017; Micceri, 2089) y no podría representar un problema ni para caracterizar a las conductas medidas ni para el modelamiento paramétrico de todo el instrumento. De acuerdo con cómo serán tratados los datos en el análisis, parametrizaciones SEM para ítems tratados como variables categóricas son efectivos con algunos estimadores como DWLS o ULS (Muthén, du Toit, Spisic, 1997), aunque también cuando son tratadas como variables continuas (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009), y son actualmente estimadores razonables de los parámetros de los ítems. Una implicación es que el tratamiento conjunto de los ítems del PROPSIT en sus dimensiones esperadas requiere estimadores que tomen en cuenta esta no normalidad predecible, y adicionalmente, considerando la extensión del escalamiento (ie., siete opciones de respuesta), pueden considerarse presunciones de continuidad de respuesta (Brown, 2015; Forero et al., 2009).

Respecto a los resultados de asociación, la sustancialidad de los resultados muestra en general una concordancia teórica de los contenidos con las variables externas elegidas; en este sentido, el patrón asociaciones monotónicas positivas y negativas detectadas en ambas áreas (factores de riesgo psicosocial y recursos psicosociales favorables) indicaron que los indicadores señalados son marcadores con significancia teórica y práctica. Aunque la magnitud predominante de las asociaciones obtenidas puede considerarse pequeña o moderada si se comparan con guías sugestivas de interpretación (e.g., Cohen, 1992), estas guías fueron construidas para aplicarse en variables continuas y no pueden ser directamente trasladadas a nuestros resultados ni para cualificar ni estimar el poder estadístico (Correll, Mellinger, McClelland, & Judd, 2020). La interpretación de los análisis de ítems con el objetivo de validación de una medida requiere adaptación de tales guías, y aplicado a nuestros resultados, la dirección negativa o positiva de las correlaciones puede ser priorizada hasta dilucidar un esquema claro para interpretar las magnitudes obtenidas. Sin embargo, aún con esta limitación, el tamaño de las correlaciones obtenidas son indicadores de satisfactoria validez de los ítems.

Una cuestión metodológica implicada en el análisis de la asociación anterior requiere atención. En el estudio se aplicó una medida general de dependencia basada en la teoría de la información (coeficiente MIC), para detectar alguna funcional entre las variables, pero predominantemente arrojaron resultados sin significancia estadística, esto es, que la asociación no presentó alguna forma funcional. Mientras, por un lado, los coeficientes MIC de asociación general no parecieron alcanzar poder estadístico para detectar alguna forma funcional de dependencia entre las variables, por otro lado, sí se detectaron asociaciones monotónicas entre las variables mediante el coeficiente de monotonicidad de Spearman, sugiriendo que la categorización ordinal puede crear fuerte entropía que impide distinguir patrones densos (i.e., bajo “ruido” o *noise*) y uniformes bivariadamente. Enfocándose en los resultados sustanciales derivados de la correlación monotónica, los ítems de factores de riesgo y recursos favorables covariaron con más fuerza con las medidas de engagement y distrés (UWES-3 y PHQ-4, respectivamente), comparado con las asociaciones obtenidas con el estrés (medida de ítem único). Respecto a los ítems que mostraron menor debilidad asociativa, o los que mostraron asociaciones contraintuitivas (demandas cognitivas-atencionales como la atención prolongada a tareas, a dos o más tareas y elevado esfuerzo mental, y demandas de carga y ritmo de trabajo como el monto y rapidez de las tareas laborales), requieren mayor exploración para discernir los factores que influenciaron en este tipo de resultados. Podemos hacer la conjetura que, la intensidad de las demandas laborales evaluadas en la muestra fue restringida a las intensidades bajas, con altas demandas más bien episódicas (Reina-Tamayo et al., 2017), de tal modo que produjeron una manejable experiencia para resolverlos y en estos casos se perciben más como retos que exigencias, lo que a su vez influye en sentimientos de maestría y otras experiencias positivas. Esto puede corroborarse en la tendencia de respuesta, donde las opciones que indican menor intensidad de estos estresores mostraron mayor densidad de respuesta. Sin embargo, otros aspectos no evaluados que pudieron moderar esto, como la deseabilidad social, o la representación sociocognitiva individual de lo que significa tener “mucho trabajo” o estar “siempre

ocupado” en el contexto peruano. Ambos y otros aspectos pudieron ser parcialmente responsables, y requieren futura exploración sobre su replicabilidad y explicación probables.

Una implicación general del presente estudio apunta a la práctica de publicaciones sobre validez de medidas de factores psicosociales, hay una aparente necesidad de espacio propio para análisis detallados en el nivel de los ítems, para respaldar conclusiones complementarias sobre la caracterización psicométrica de las medidas estudiadas, obtenidas con la determinación de la estructura factorial y su confiabilidad, que por cuestiones de espacio y énfasis que se desea en la exploración, no siempre es conveniente presentarse dentro de las mismas publicaciones.

En presente estudio tiene algunas limitaciones: primero, el tamaño muestral pudo ser condicionante para la distribución de respuestas en las opciones, y que aumentó el error tipo I en la identificación de categorías no utilizables, así que un tamaño muestral con poder estadístico suficiente para superar la frecuencia mínimo de 10 respuestas es altamente recomendado. Segundo, la representatividad de la muestra respecto a las ocupaciones muestreadas no está garantizada, debido que el diseño del estudio orientado a maximizar la heterogeneidad de las ocupaciones. Tercero, y asociado a la anterior, esta heterogeneidad pudo también ser determinante para la emergencia de opciones de respuesta subutilizadas, por ejemplo, cuando algunos factores de riesgo psicosocial están fuertemente vinculados con la intensidad de la misma, como el esfuerzo físico. Cuarto, otros aspectos de la calidad del sistema ordinal del PROPSIT no fueron evaluados y que corresponden a un conjunto de indicadores de calidad (Linacre, 2002), como la distancia entre los umbrales, la monoticidad de la relación ítem-puntaje, la progresión monótonica de las opciones de respuesta, etc. Sin embargo, esto requiere la definición de la dimensionalidad de cada ítem como presunción fundamental, y ajustar a un modelo paramétrico, como el modelo de crédito parcial. En esta etapa de la investigación, la atención fue puesta en las propiedades de los ítems como elementos individuales, vinculados por construcción teórica. La inspección de la dimensionalidad rigurosamente abordado requiere una extensión de la

mitad de este manuscrito o más, y para evitar falsos positivos en la identificación de la dimensionalidad, esta meta queda como meta futura de investigación. Quinto, las medidas basadas en ítems únicos tienen ventajas y limitaciones, y en relación con el ítem único de estrés (UIE; Elo et al., 2003), se requiere discernir entre la falta de validez de este ítem para el contexto peruano en general, la ausencia de poder estadístico, o su falta de sensibilidad para los contenidos muestreados en el PROPSIT. Finalmente, la deseabilidad social no fue incluida como otra variable externa, y es altamente recomendado diseñar su inclusión para evaluar el posible impacto de este estilo de respuesta.

El estudio concluye esencialmente en dos cosas: Primero, que las relaciones entre los ítems del PROPSIT y medidas externas mantienen relaciones teóricas consistentes y entre bajas y moderadas para el contenido definido para los recursos favorables laborales, mientras que los factores de riesgo psicosocial existieron algunas discrepancias que pueden provenir de las idiosincrasia muestral o de sesgos de deseabilidad social; segundo, con base en esta experiencia, es claro que el estudio de la validez de constructo de los ítems no solo debería incorporar la fuerza de la relación ítem-constructo en la métrica de los resultados del modelamiento factorial, sino también la asociación con criterios de constructo o conductuales externos, los cuales tienen consecuencias para la validez de contenido así como para la interpretación de los puntajes que se obtendrán, logrando así una balance más integral de las propiedades psicométricas de diversas medidas.

3.4. Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST

3.4.1. Introducción

El objetivo de este trabajo fue investigar la *respuesta con esfuerzo insuficiente* (REI) en los datos recolectados, y que sirve como evaluación precursora de la calidad de los datos para realizar el examen de la dimensionalidad y de las relaciones sustantivas del estudio de la adaptación del PROPSIT. La respuesta descuidada (*careless*; Meade & Craig, 2012) o con esfuerzo insuficiente (*insufficient effort*; Curran, 2016), llanamente se refiere a la atención disminuida puesta al contenido del ítem. Este problema introduce respuestas inválidas porque no se asocian al contenido de los ítems, y parece omnipresente en las evaluaciones basadas en cuestionarios. La prevalencia obtenida en diferentes estudios sobre este problema, varía en la mínima y máxima cantidad estimada; por ejemplo, entre 3% y 40% (Johnson, 2005; Kurtz & Parrish, 2001; Maniaci, Rogge, & 2014; Oppenheimer, Meyvis, & Davidenko, 2009). Esta variación esta a su vez sujeta a variaciones de muestreo, así como a variaciones en los métodos para identificar REI, a las condiciones de instrucción de llenado, las características del instrumento, y las variaciones locales del contexto de la aplicación de las encuestas; por lo tanto, no son estimaciones estáticas, y lo que se obtiene de la literatura con referencias provisionales de valor heurístico.

La importancia de detectar REI en los datos está vinculada con el conocimiento de su prevalencia, su tratamiento y su prevención en la metodología de encuestas, pero también en la calidad de los resultados. Por ejemplo, en medidas unidimensionales, una prevalencia del 10% a 20% pueden distorsionar el modelamiento confirmatorio (Woods, 2006), y conducir a decisiones incorrectas sobre la dimensionalidad y configuración factorial de los instrumentos. Entre los parámetros específicos que son afectados por REI están las cargas factoriales y correlaciones factoriales, que se inflan cuando los participantes con REI no son removidos (Kam, 2019; McGonagle et al., 2016). Por lo tanto, una consecuencia de la detección de REI es la obtención de estimaciones más realistas, aunque

decrementadas (McGonagle et al., 2016) debido que disminuye la varianza espuria en los datos. El efecto de REI sobre los resultados sustantivos o psicométricos puede entenderse como una cadena de efectos que en común deterioran las estimaciones (Schroeders, Schmidt, & Gnambs, 2022).

Efectivamente, hay evidencia que señala el deterioro de las estimaciones de validez y confiabilidad, así como el aumento del error Tipo I y Tipo II (Arias et al., 2020; Burchett et al., 2016; Holtzman & Donnellan, 2017).

Debido que la evaluación de los factores psicosociales preferentemente recurre al uso de medidas de autoinforme, con metodología lápiz-papel o en plataforma online, las implicaciones de REI también deberían estar presentes también en estas evaluaciones. En los estudios del área laboral en general, y en los factores psicosociales del trabajo en Latinoamérica en particular, apenas se ha introducido este asunto. Aunque se espera que el control de sesgos sea parte de las estrategias incorporadas en el diseño de investigación (Huang, Curran, Keeney, Poposki, & DeShon, 2012), los estudios de identificación de REI y de sus efectos sobre las estimaciones de confiabilidad, validez y clasificaciones de los grupos de trabajadores, aun parece un conocimiento incierto, y que poco ha llamado la atención de los evaluadores e investigadores en el estudio de factores psicosociales del trabajo. El llamado a la comprensión de este problema, su detección y prevención ha sido enfatizado hace pocos años (McGonagle, Huang, & Walsh, 2016), pero posiblemente la presunción de control durante la recolección de datos, así como creer que sus efectos son inocuos, parece prevalecer.

Los estudios de prevalencia, correlatos y efectos de REI se reportan en mayor número en participantes anglosajones, y apenas hay estudios en población hispana, y específicamente en trabajadores en quienes uno de los focos de estudio importantes apunta a los factores psicosociales del trabajo. Por ejemplo, el meta-análisis de Dodou y de Winter's (2014) se observa una predominancia de estudios con participantes no hispanos, y en diferentes áreas de actividad. Aunque los resultados concretos pueden generalizarse a estudios hispanos si se considera un ángulo ético, el parecido o

diferencia que podría encontrarse en trabajadores hispanos, o peruanos en particular, es una pregunta no respondida aún. De acuerdo a esto, se investigó la ocurrencia de REI y sus efectos en la exploración de la dimensionalidad del PROPSIT.

3.4.2. Método

Participantes

La población de referencia fueron dos grupos de trabajadores: un grupo heterogéneo de diferentes puestos de trabajo y carreras, en diferentes organizaciones laborales, predominantemente de Lima Metropolitana, y un grupo de personal docente de una institución de educación regular, en Lima Metropolitana. El primer grupo fue la muestra del Objetivo 2b-2 (Examinar la claridad, relevancia y validez con criterios externos del contenido del PROPSIT), y en que se evaluó la validez de contenido de los ítems con criterios externos. En el grupo de docentes, la disponibilidad de la institución para participar en el estudio fue el criterio de elección. En la Tabla 21 aparece la distribución de participantes de la muestra completa (trabajadores en general y docentes), y la muestra luego de remover los casos detectados con REI (ver más adelante el procedimiento efectuado) o muestra “limpia”.

Tabla 21

Características demográficas y laborales de los partes (antes y después de la limpieza de datos)

	Muestra completa (N = 359)		Muestra "limpia" (N = 305)	
	N	%	N	%
Base de datos				
General	201	56.0	184	60.3
Educación	158	44.0	121	39.7
Sexo				
Hombre	171	47.6	146	47.9
Mujer	187	52.1	159	52.1
Perdido	1	.3		
Lugar de nacimiento				
Lima	159	44.3	147	48.2
Fuera de Lima	186	51.8	148	48.5
Perdido	14	3.9	10	3.3
Instrucción				
Primaria terminada	2	.6		
Secundaria terminada	40	11.1	35	11.5
Bachiller terminado/ Carrera técnica terminada	158	44.0	138	45.2
Licenciatura/ Universidad terminado	89	24.8	73	23.9
Posgrado terminado	56	15.6	50	16.4
Perdido	14	3.9	9	3.0
<i>Clasificación de la profesión</i>				
Universitario	288	80.2	246	80.7
Técnico	47	13.1	41	13.4
No universitario/ técnico	19	5.3	18	5.9
Perdido	5	1.4		
Contrato				
Planta/Definitivo	186	51.8	155	50.8
Eventual/Temporal	148	41.2	131	43.0
Perdido	25	7.0	19	6.2
Convivencia				
Solo	18	5.0	14	4.6
Solo con mascota	2	.6	2	.7
Con pareja o familiares	315	87.7	270	88.5
Con amigos	3	.8	3	1.0
Otro	19	5.3	15	4.9
Perdido	3	.9	1	.3
Clasificación de áreas académicas				
Ciencias de la salud	6	1.7	5	1.6
Ciencias básicas	3	.8	2	.7
Ingeniería	23	6.4	21	6.9
Ciencias económicas y de la gestión	69	19.2	63	20.7
Humanidades, ciencias jurídicas y sociales	193	53.8	155	50.8
No pertenece/ no aplica	64	17.8	59	19.3
Perdido	1	.3		
Clasificación de trabajos				
Directores y gerentes	3	.8	1	.3
Profesionales científicos e intelectuales	214	59.6	179	58.7
Técnicos y profesionales de nivel medio	55	15.3	55	18.0
Personal de apoyo administrativo	27	7.5	27	8.9
Trabajadores de servicios, vendedores de comercios y mercados	9	2.5	10	3.3
Oficiales, operarios, artesanos y artes mecánicas y de otros oficios	28	7.8	28	9.2
Operadores de instalaciones, máquinas y ensambladores	4	1.1	4	1.3
Ocupaciones elementales	1	.3	1	.3
Perdido	18	5.0		

Instrumento

Escala de Procesos Psicosociales del Trabajo (Juárez-García, & Flores-Jiménez, 2020). Se utilizó el mismo instrumento del estudio para el objetivo Objetivo 2b-2.

Procedimiento

Recolección de datos. En el grupo heterogéneo de trabajadores, la descripción del proceso de aplicación aparece en la sección participantes del Objetivo 2b-2. Por otro lado, el grupo docentes fue expuesto a aplicación presencial, y monitoreado por una docente, colaboradora del investigador para lograr el acceso. La docente recibió instrucciones del investigador para la aplicación estandarizada de la encuesta, y se enfatizó la firma la aceptación voluntaria para participar, la libertad para continuar durante el llenado de la encuesta, el anonimato de respuesta, la confidencialidad de los resultados individuales y la atención que deberían dedicar a las instrucciones y contenido de la encuesta. Se mantuvo el mismo orden de presentación de los instrumentos en ambas muestras.

Análisis de valores perdidos. Debido a la presencia de valores perdidos en la base de datos (ie, *missing values*), primero se verificó si los datos perdidos pudieron haber ocurrido por un mecanismo aleatorio (*missing completely at random*, MCAR; Little, 1998), esto es, que no tuvieran asociación con las variables (ie, ítems). Para ello se implementó la prueba global basada en la distribución χ^2 (Little, 1988), cuya hipótesis nula es que los valores perdidos son completamente aleatorios. Este análisis utilizó el script R *mcAR* (Stemmler, 2020). También se usó una prueba basada en regresión de los ítems sin datos perdidos sobre cada ítem independiente con datos perdidos (*regression-based*, RB; Rouzinov, & Berchtold, 2020). Una vez que se constató que los valores perdidos pueden haberse generado aleatoriamente, el tratamiento de los datos perdidos fue hecho mediante el método de imputación de dos vías (*two-way imputation*, TW; Bernaards, & Sijtsma, 2000), un método basado en el modelo ANOVA de dos vías, y proveniente de la media de respuestas a los ítems y la media de respuestas del

individuo. Está representado por la fórmula: $TW_{ij} = PM_i + IM_j - OM$, donde PM_i es la media de respuestas al ítem de la persona i , IM_j es la media de respuestas al ítem j , y OM es la media total de todos los ítems. El método TW produce menos discrepancias con parámetros psicométricos en datos completos, comparado con otros métodos aplicados para datos de cuestionarios (Bernaards, & Sijtsma, 2000; van Ginkel, van der Ark, & Sijtsma, 2007).

Análisis de sesgos de respuesta. Se implementó la detección de dos patrones de respuestas asociados al modo descuidado de responder a ítems de auto-informes: la respuesta aleatoria y la secuencia larga de respuestas iguales (Curran, 2016; Meade, & Craig, 2012). La respuesta aleatoria parece estar vinculada con mayor variabilidad individual de respuesta (Dunn, Heggstad, Shanock, & Theilgard, 2018; Finch, 2012). Por lo tanto, se detectó la presencia de valores extremos multivariados (*outliers*) mediante la *distancia Mahalanobis* (D^2 ; Mahalanobis, 1936); este puntaje se calcula para cada participante, y se lo compara con los cuantiles de la distribución χ^2 , con el número de grados de libertad igual al número de ítems analizados (41 ítems del PROPSIT). El criterio para identificar a los casos extremos en el nivel 99% de confianza, fue $\chi^2 > 64.95$. En contraste con otros métodos de detección de outliers multivariados, D^2 no produce una remoción grande de datos (Finch, 2012). Por otro lado, la secuencia larga de respuestas iguales (*longstring*; Curran, 2016; Meade, & Craig, 2012) es un patrón de respuesta también vinculado con la respuesta descuidada, y expresado como poca variabilidad de respuesta. Este patrón se identifica cuando el número de respuestas idénticas es mayor a la mitad del número de variables analizadas (Curran, 2016); por lo tanto, el punto de corte para este estudio fue $\text{longstring} > 20$.

Modelamiento de la estructura interna. Para evaluar el efecto de REI sobre el modelamiento del PROPSIT, se primero se hizo el ajuste del modelo de medición del PROPSIT en los datos recogidos. Debido que existe una estructura predefinida de configuración de los ítems del PROPSIT (ie, la ubicación de los ítems en sus dimensiones), el estudio de las dimensiones de los ítems del

PROPSIT se efectuó con el procedimiento de *análisis factorial semi-confirmatorio* (Browne, 1972; McDonald, 2005), en cada sección por separado (e.g., riesgo psicosocial y recursos favorables) para examinar con detalles cada una, y el formato de presentación de los ítems sugirió una sección distinta de contenido entre los ítems de riesgo psicosocial y recursos favorables. Este efectuó una rotación de tipo Procusto hacia una matriz definida a priori (*matriz target*), y en que se describe la ubicación de los ítems en los factores del modelo de medición. Esta matriz es la hipótesis sobre la estructura interna del instrumento. Para efectuarlo, esta configuración se basó en la distribución teórica de los ítems del PROPSIT, que aparece representada en la Tabla 22 y Tabla 23 para los factores de riesgo psicosocial (FPST) y recursos favorables (RF). El código “9” indica la ubicación esperada del ítem en su factor, y cuya carga factorial se estimará libremente, y “0” indica la cantidad hacia donde será rotada la solución factorial, acercándose a este valor. Para FRPS, los factores en este modelo fueron demandas de *carga y ritmo de trabajo* (F1), de *responsabilidad y peligrosidad* (F2), de *jornada, turno y horario* (F3), *cognitivas – atencionales* (F4), *emocionales* (F5), de *esfuerzo físico*, y *acoso psicológico* (F7). Para RF: fueron: *recompensas y desarrollo profesional* (F1), *control laboral y contenido de la tarea* (F2), *recursos para realizar el trabajo* (F3), y *clima y apoyo social laboral* (F4).

Tabla 22

Factores de Riesgo Psicosocial: Configuración inicial (matriz target) para la primera iteración

	F1	F2	F3	F4	F5	F6	F7
1. Carga de trabajo	9	0	0	0	0	0	0
2. Trabajar rápido	9	0	0	0	0	0	0
3. Interrupciones para tareas a tiempo	0	9	0	0	0	0	0
4. Tomar acciones importantes	0	9	0	0	0	0	0
5. Peligros	0	9	0	0	0	0	0
6. Horas extras, horarios prolongados	0	0	9	0	0	0	0
7. Rotación/cambio de turno	0	0	9	0	0	0	0
8. Jornadas en turno nocturno	0	0	9	0	0	0	0
9. Atención prolongada en tareas	0	0	0	9	0	0	0
10. Atender dos o más tareas	0	0	0	9	0	0	0
11. Esfuerzo mental	0	0	0	9	0	0	0
12. Emociones negativas otras personas	0	0	0	0	9	0	0
13. Trato con personas desgastante	0	0	0	0	9	0	0
14. Mostrar emociones diferentes	0	0	0	0	9	0	0
15. Mucho esfuerzo físico	0	0	0	0	0	9	0
16. Posiciones incómodas	0	0	0	0	0	9	0
17. Condiciones ambientales adversas	0	0	0	0	0	9	0
18. Maltrato psicológico (superiores)	0	0	0	0	0	0	9
19. Maltrato psicológico (compañeros)	0	0	0	0	0	0	9
20. Demasiado control	0	0	0	0	0	0	9
21. Reglas y normas exageradas	0	0	0	0	0	0	9
22. Retroalimentación supervisores (R)	0	0	0	0	0	0	-9

Nota. 9: Código que indica la ubicación del ítem, en que la carga factorial será estimada libremente. 0: valor hacia donde será aproximado la carga factorial durante la rotación Procusto. R: ítem que debe ser recodificado para su interpretación. F1: carga y ritmo de trabajo. F2: responsabilidad y peligrosidad; F3: jornada, turno y horario. F4: cognitivas – atencionales. F5: emocionales. F6: esfuerzo físico. F7: acoso psicológico.

Tabla 23

Factores de Recursos Favorables: Configuración inicial (matriz target) para la primera iteración

	F1	F2	F3	F4
23. Trabajo justo y equitativa.	9	0	0	0
24. Salario motivante	9	0	0	0
25. Trabajo apreciado y reconocido	9	0	0	0
26. Sin oportunidades profesionales (R)	-9	0	0	0
27. Perder este empleo (R)	-9	0	0	0
28. Trabajo agradable y gratificante	9	0	0	0
29. Tareas benefician a personas y sociedad	9	0	0	0
30. Libre en decidir el trabajo	0	9	0	0
31. Utilizo habilidades	0	9	0	0
32. Desarrollo de capacidades	0	9	0	0
33. Actividades muy variadas	0	9	0	0
34. Funciones y tareas muy claras	0	9	0	0
35. Materiales necesarios y adecuados	0	0	9	0
36. Capacitaciones necesarias	0	0	9	0
37. Compañeros apoyan	0	0	0	9
38. Superiores apoyan	0	0	0	9
39. Clima de unión/colaboración	0	0	0	9
40. Coincidencia de valores con organización	0	0	0	9
41. Coincidencia de valores con compañeros	0	0	0	9

Nota. 9: Código que indica la ubicación del ítem, en que la carga factorial será estimada libremente. 0: valor hacia donde será aproximado la carga factorial durante la rotación Procusto. R: ítem que debe ser recodificado para su interpretación. F1: recompensas y desarrollo profesional. F2: Control laboral y contenido de la tarea. F3: recursos para realizar el trabajo. F4: clima y apoyo social laboral.

El estimador para la extracción de los factores fue *mínimos cuadrados no ponderados* (*unweighted least squares*, ULS; McDonald, 1982), un método que produce resultados aproximadamente exactos cuando existe moderada o fuerte ausencia de normalidad y asimetría distribucional de los datos, cuando el número de factores es mayor a 3 (como ocurre en los presentes datos; Morata-Ramirez, Holgado-Tello, Barbero-Garcia, & Mendez, 2015; Ximénez, & García, 2005), y los indicadores son ordinales (Li, 2016; Muthén, 1993). La evaluación del ajuste implementó estimaciones robustas, mediante una corrección a la media y varianza del estadístico χ^2 de ajuste

(ULSMV; Muthén, 1993; Satorra & Bentler, 1994). Considerando el número de 7 opciones de respuesta de los ítems del PROPSIT, el estimador ULS se aplicó a las correlaciones Pearson entre los ítems. Para evaluar el grado de ajuste de los modelos a los datos, se usó el *coeficiente de congruencia* (*C*; Tucker, 1951); este coeficiente varía entre -1.0 y 1.0, y cuantifica el grado de similaridad numérica entre el resultado del análisis factorial y la matriz *target* a priori establecida. Este coeficiente indica una aceptable ($> .85$), o alta similaridad o igualdad práctica ($> .95$) entre las cargas factoriales comparadas (Lorenzo-Seva, & ten Berge, 2006). En análisis factorial se hizo mediante el programa *Factor* (Ferrando, & Lorenzo-Seva, 2018).

Efectos de REI sobre el modelamiento del PROPSIT. Para evaluar el efecto de los sesgos de respuesta detectados sobre el modelo de medición del PROPSIT, se obtuvieron y compararon varios indicadores de la dimensionalidad y estructura interna entre la muestra completa (“sucía” o “no limpia”) y la muestra sin las posibles REI (“limpia”). Estos indicadores comparados fueron de: a) dimensionalidad, con la búsqueda del número de dimensiones latentes, mediante el *Empirical Kaiser Criterion* (*EKC*; Braeken & van Assen, 2017); b) de estabilidad de los factores, mediante el coeficiente de *determinabilidad factorial* (Ferrando & Lorenzo-Seva, 2018); c) de ajuste, mediante el coeficiente de congruencia (*C*); y, d) de confiabilidad, basado en los puntajes factoriales del modelo ajustado, mediante *Overall Reliability of fully-Informative prior Oblique N-EAP scores* (ORION; Ferrando, & Lorenzo-Seva, 2018). La diferencia entre cada parámetro psicométrico anterior (ie., θ) fue estimada mediante el cambio porcentual: $100 * (\theta_{limpia} - \theta_{sucia}) / \theta_{limpia}$.

3.4.3. Resultados

Análisis de valores perdidos

Luego de verificar la existencia de la misma codificación y número de variables de interés, la primera evaluación de los datos arrojó que 29 sujetos no respondieron a una cantidad inviable para el

análisis principal, distribuidos en 23 participantes sin respuesta para 16 ítems (39%), 4 para 17 ítems (41.4%), 1 para 18 ítems (43.9%), y 1 para 19 ítems (43.9%). Estos participantes fueron removidos de la base de datos para evitar resultados espurios producidos por la imputación de datos posterior. Por otro lado, 1, 2 y 3 ítems no fueron respondidos por 43, 4 y 2 participantes, respectivamente; estos 49 participantes permanecieron en la base de datos para la evaluación de MCAR. MCAR fue examinado para el área Riesgos Psicosociales (FRP, 22 ítems) y Recursos Favorables (RF, 19 ítems) por separado. Se halló resultados estadísticamente no significativos en FRSP (MCAR – $\chi^2 = 430.868$, $gl = 393$, $p = .098$), pero en RF la evidencia fue débil para aceptar la hipótesis de aleatoriedad de los valores perdidos (MCAR – $\chi^2 = 407.198$, $gl = 215$, $p < .001$). Sin embargo, la prueba RB para cada una de los ítems arrojó que todos los ítems con valores perdidos no mostraron relación con los ítems con datos completos, asumiendo un patrón MCAR en cada ítem. Debido que, para cada ítem el patrón pareció MCAR con menos de 1% de datos perdidos, se hizo la imputación *two-way* asumiendo MCAR, y se continuó con los siguientes análisis.

Prevalencia de respuestas descuidadas

Respecto a las REI, se detectaron 48 puntajes extremos (13.3%), con puntajes D^2 mayores a 64.95 ($M = 40.88$, $min = 7.56$; $max = 130.41$); y respecto al indicador de la secuencia de respuestas iguales (*longstring*), se detectaron 6 participantes (1.6%) con respuestas secuenciales iguales entre 22 y 41 ítems (Figura 13). La prevalencia conjunta de participantes con posibles REI fue de 14.9% ($n = 54$), un porcentaje que ligeramente supera la prevalencia reportada en la literatura (Arias et al., 2020). Los casos con REI se distribuyeron respectivamente en 23.4% y 8.4% de la muestra educadores y de trabajadores en general. Luego de remover estos 54 participantes, la muestra efectiva total para el análisis de la estructura interna fue 305 participantes.

Efectos sobre el modelamiento del PROPSIT

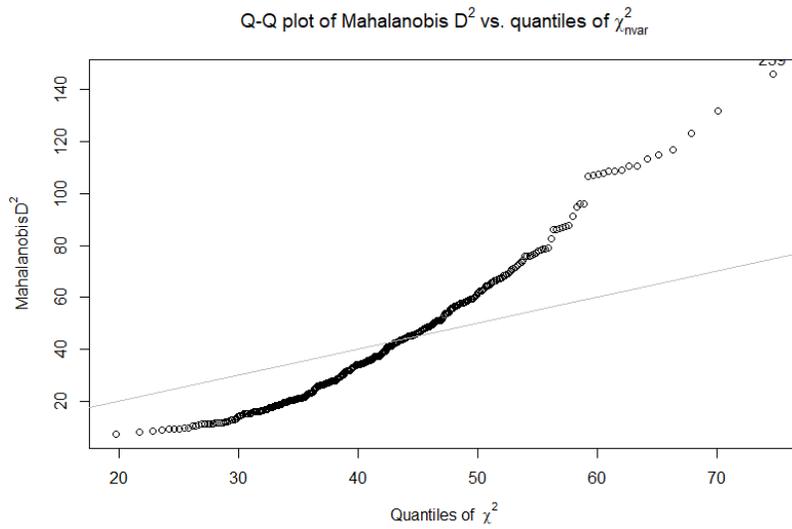
Matriz correlacional. En los ítems de Factores de Riesgo Psicosocial (FRPS), la matriz correlacional permaneció adecuada para la factorización, después ($KMO = .832$, $Bartlett-\chi^2 = 3405.9$, $p < .01$) y antes de remover las repuestas REI ($KMO = .850$, $Bartlett-\chi^2 = 3539.1$, $p < .01$), pero con leves cambios.

Dimensionalidad. En Factores de Riesgo Psicosocial (FRPS), la exploración del número de factores con el criterio empírico de Kaiser (EKC; Figura 14) halló autovalores mayores a 1.0 en la muestra limpia (6.34, 3.49, 1.79, 1.64 y 1.01), pero que apenas se diferenciaron de los autovalores en la muestra completa (6.51, 3.18, 1.71, 1.61, 1.12); en ambos grupos, el EKC sugiere cinco factores interpretables. En Recursos Favorables (RF), se obtuvo similar patrón de autovalores mayores a 1.0 (muestra completa: 8.85, 1.86, 1.35 y 1.00; muestra limpia: 10.03, 1.73, 1.27 y 1.04); el número indicado por el método EKC en ambas muestras fue cuatro factores.

Figura 13

Detección de respuestas con probable insuficiente esfuerzo

Outliers multivariados (distancia Mahalanobis)



Boxplot de casos con longstring detectados

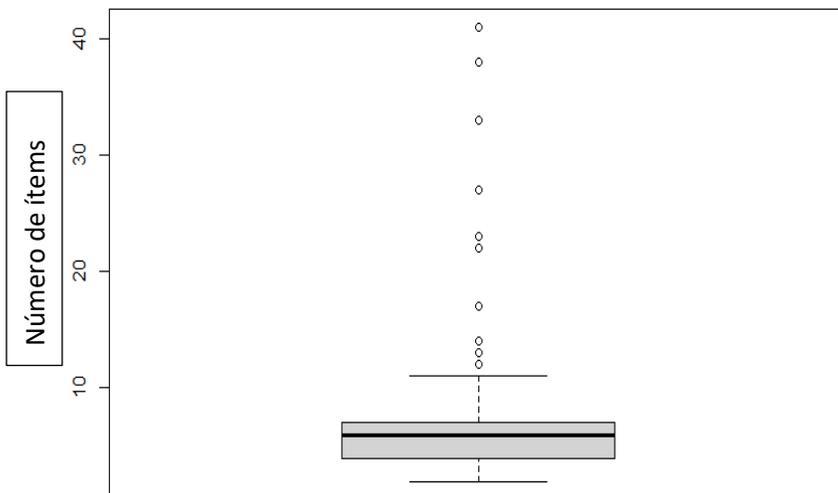
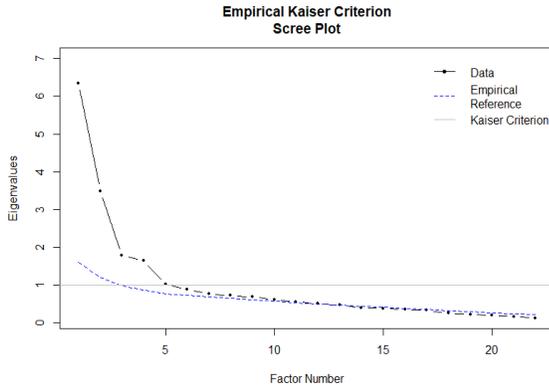


Figura 14

Criterio empírico de Kaiser (EKC)

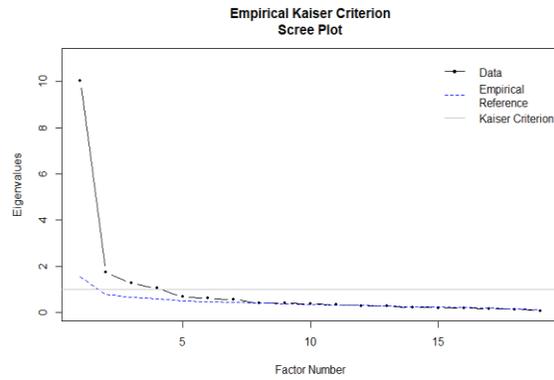
Factores Psicosociales del Trabajo

Muestra "limpia"

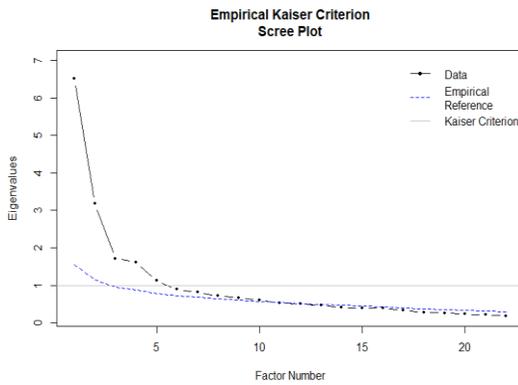


Factores Recursos Favorables

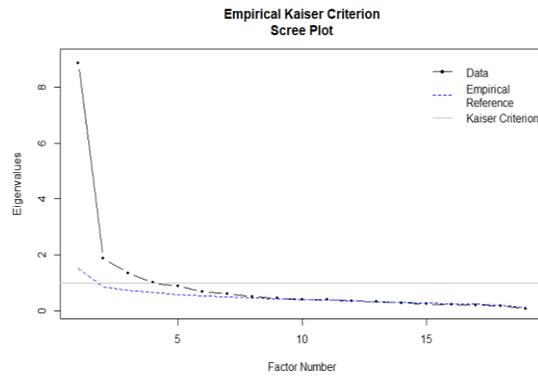
Muestra "limpia"



Muestra completa



Muestra completa



Calidad de los resultados factoriales. En la Tabla 24 para ambos grupos (completa y limpia), se presenta la varianza retenida, el índice de determinabilidad factorial (IDF) y la confiabilidad de los puntajes factoriales por el método ORION. En el cambio porcentual, indicado en la misma Tabla 24 para los factores, señala que hubo una predominante mejora en estos indicadores de calidad, excepto

para F1 de FPST, y F1 de RF. El cambio porcentual en las cargas factoriales (Tabla 25) indicó un patrón inconsistente en la dirección (decremento o incremento) y el tamaño (entre -136.1% y 33.3% en los ítems de factores de riesgo psicosocial; entre -425.0% y 40.0% en los factores de recursos favorables).

Tabla 24

Calidad comparativa de la solución factorial (muestra completa vs muestra limpia)

	Muestra completa (n = 358)			Muestra limpia (n = 305)			Cambio porcentual		
	Varianza retenida	IDF	ORION	Varianza retenida	IDF	ORION	Varianza retenida	IDF	ORION
FRPS									
F1	1.451	0.92	.85	0.86	.87	.76	-68.7%	-5.7%	-11.8%
F2	1.282	0.87	.75	1.43	.90	.81	10.3%	3.3%	7.4%
F3	1.374	0.88	.77	1.17	.90	.82	-17.4%	2.2%	6.1%
F4	2.582	0.94	.88	3.51	.95	.92	26.4%	1.1%	4.3%
F5	1.679	0.89	.80	1.76	.90	.82	4.6%	1.1%	2.4%
F6	1.871	0.94	.89	1.92	.94	.88	2.6%	0.0%	-1.1%
F7	2.448	0.94	.89	2.65	.97	.94	7.6%	3.1%	5.3%
RF									
F1	1.92	.90	.81	1.08	.836	.69	-77.8%	-7.7%	-17.4%
F2	4.14	.97	.95	5.04	.985	.97	17.9%	1.5%	2.1%
F3	1.37	.90	.81	2.33	.935	.87	41.2%	3.7%	6.9%
F4	3.67	.95	.91	4.15	.969	.93	11.6%	2.0%	2.2%

Nota. FRPS: Factores de Riesgo Psicosocial. RF: Recursos Favorables. IDF: índice de determinabilidad factorial. ORION: confiabilidad basada en los puntajes factoriales. Cambio porcentual: $100 * (\theta_{limpia} - \theta_{sucia}) / \theta_{limpia}$.

Tabla 25

Coefficientes de congruencia para los ítems y cambio porcentual

Ítems	Factores de Riesgo Psicosocial (FRPS)			Ítems	Factores de recursos Favorables (RF)		
	Completa	Limpia	% _{dif}		Completa	Limpia	% _{dif}
PROP1	.621	.437	-42.1%	PROP23	.86	.33	-16.6%
PROP2	.94	.522	-8.1%	PROP24	.84	.16	-425.0%
PROP3	.441	.661	33.3%	PROP25	.87	.19	-357.9%
PROP4	.71	.556	-27.7%	PROP26	.91	.96	5.2%
PROP5	.758	.861	12.0%	PROP27	.86	.93	7.5%
PROP6	.654	.277	-136.1%	PROP28	.52	.23	-126.1%
PROP7	.849	.854	.6%	PROP29	.09	-.03	40.0%
PROP8	.729	.812	1.2%	PROP30	.94	.98	4.1%
PROP9	.969	.968	-.1%	PROP31	.99	.99	.0%
PROP10	.883	.972	9.2%	PROP32	.99	.99	.0%
PROP11	.9	.894	-.7%	PROP33	.96	.97	1.0%
PROP12	.934	.966	3.3%	PROP34	.97	.99	2.0%
PROP13	.949	.897	-5.8%	PROP35	.93	.96	3.1%
PROP14	.966	.967	.1%	PROP36	.97	.95	-2.1%
PROP15	.984	.984	.0%	PROP37	.91	.95	4.2%
PROP16	.965	.994	2.9%	PROP38	.98	.98	.0%
PROP17	.842	.797	-5.6%	PROP39	.97	.99	2.0%
PROP18	.825	.85	2.9%	PROP40	.99	.99	.0%
PROP19	.858	.882	2.7%	PROP41	.97	.98	1.0%
PROP20	.922	.895	-3.0%	-	-	-	-
PROP21	.87	.91	4.4%	-	-	-	-
PROP22	.271	.229	-18.3%	-	-	-	-
Factores				Factores			
	Completa	Limpia	% _{dif}		Completa	Limpia	% _{dif}
F1	.70	.57	-22.8%	F1	.86	.70	-22.9%
F2	.68	.78	12.8%	F2	.90	.90	.0%
F3	.65	.62	-4.8%	F3	.85	.65	-3.8%
F4	.80	.68	-17.6%	F4	.94	.93	-1.1%
F5	.90	.92	2.2%	-	-	-	-
F6	.94	.93	-1.1%	-	-	-	-
F7	.89	.89	.0%	-	-	-	-
Total	.811	.783	-3.6%	Total	.87	.78	-11.5

Nota. Muestra completa: n = 358. Muestra limpia: 305. %_{dif}: Cambio porcentual definido como $100 * (\theta_{limpia} - \theta_{sucia}) / \theta_{limpia}$.

Factores de Riesgo Psicosocial:

F1: carga y ritmo de trabajo. F2: responsabilidad y peligrosidad; F3: jornada, turno y horario. F4: cognitivas – atencionales. F5: emocionales. F6: esfuerzo físico. F7: acoso psicológico.

Factores Recursos Favorables:

F1: recompensas y desarrollo profesional. F2: Control laboral y contenido de la tarea. F3: recursos para realizar el trabajo. F4: clima y apoyo social laboral.

3.4.4. Discusión

El presente objetivo apuntó a la detección de REI, hacia su prevalencia y su efecto sobre la adaptación del PROPSIT en trabajadores peruanos. Comparado con el amplio rango de prevalencia reportado en la literatura, que puede expresarse en un monto tan pequeño alrededor del 3% (Johnson; Maniaci) hasta alrededor del 40% (Oppenheimer), la prevalencia hallada se encuentra dentro de este rango, y podría ser establecida como una frecuencia de REI esperada. Esto no sugiere que esta prevalencia es adecuada ni óptima, sino que el monto hallado no se desvía de tendencia en general de REI. Con una amplia frecuencia de REI en la literatura, parece más probable que la prevalencia obtenida se ubique en este rango. Debido que un aparente pequeño porcentaje de casos REI puede tener efectos serios sobre una cadena de resultados (Arias et al., 2020), y que se ha demostrado que entre 10% y 20% de casos con REI puede distorsionar el modelamiento confirmatorio (Woods, 2006), es posible pensar que la prevalencia hallada puede tener un efecto significativo sobre el modelamiento de la estructura del PROPSIT

En la identificación del número de dimensiones, en FRPS y RF no se hallaron diferencias sustanciales entre la muestra completa y limpia. Este parámetro no parece ser sensible al monto de REI detectado en los datos, pero también es probable que una posible gran distorsión en la dimensionalidad de una muestra (eg., en los participantes docentes) sea balanceada con distorsiones triviales en la otra muestra. Con una muestra más grande o muestras simuladas, esto se puede explorar más a fondo.

En cuanto a la prevalencia, la mayor proporción de participantes identificados con REI provinieron de la muestra de trabajadores docentes. Un posible motivo es que, la diferencia añadida entre este grupo y los trabajadores en general fue el contexto físico de aplicación, porque mientras los docentes llenaron la encuesta en sus lugares de trabajo, el otro grupo lo hizo en lugares que ellos eligieron (eg, hogares, puestos de trabajo, etc.). Esta diferencia puede indicar la interacción entre las

condiciones del ambiente y el interés y cuidado aplicados a responder contenidos temáticos sobre el trabajo.

Por otro lado, la presencia de ítems recodificables puede tener implicaciones para REI, debido al descuido involucrado en las respuestas genera respuestas no válidas al contenido de este tipo de ítems (Woods, 2006; Schroeders et al., 2022). Entre los resultados obtenidos aquí, los ítems recodificables (ítem 22 de FRPS; ítems 26 y 27 de RF) formaron parte de los ítems con mayores cambios porcentuales en su configuración, y su presencia en el PROPSIT pueden generar consistentes efectos de REI. Es probable que la eliminación de estos ítems durante el análisis de futuros estudios de la dimensionalidad del PROPSIT sea recomendada, tal como fueron retirados de la versión final de PROPSIT (Juárez-García & Flores-Jiménez, 2020).

Los indicadores de calidad elegidos para representar la fuerza psicométrica de los factores sirven para identificar la estabilidad de los constructos (Ferrando & Lorenzo-Seva, 2018). En este sentido, estos indicadores mejoraron en casi todos los factores de FRSP y RF, indicando que la remoción de participantes con potenciales respuestas por insuficiente esfuerzo fue efectiva. De acuerdo a la literatura (Arias et al., 2020; Holtzman & Donnellan, 2017; Kam, 2019; Kurst & Parrish, 2001; Maniaci & Rogge, 2014), no remover estos casos sospechosos ocasiona un rango de efectos sobre indicadores de validez y confiabilidad de los instrumentos, así como en indicadores de relación de variables. El mecanismo estadístico de este problema puede entenderse como el efecto de la excesiva variabilidad o la excesiva homogeneidad de las respuestas.

Posiblemente, todas las respuestas detectadas con posible REI no son necesariamente inválidas, sino que pueden aún ser útiles, pero debajo de lo óptimo (Antoun, Conrad, Couper, & West, 2019). No es claro si lo óptimo puede ser aceptado cuando está comprometido la validez de constructo en una etapa de construcción y adaptación de un instrumento. Como conclusión de este estudio, se halló la REI afectó a la configuración de la dimensionalidad del PROPSIT, con una prevalencia posiblemente

moderada en la muestra estudiada. Los análisis para los siguientes objetivos del presente proyecto deben requerir remover estos casos detectados para mejorar la calidad de los datos y garantizar resultados internamente válidos.

Una implicación final de este estudio son las decisiones que se deben tomar con los casos detectados por posible esfuerzo insuficiente, y la utilidad de incluir ítems de codificación opuesta. En la primera situación, los casos identificados pueden ser nuevamente evaluados, y los participantes ser informados del posible problema; y para el proceso de evaluación, la reducción de este problema puede involucrar monitorear la aplicación o estimular la participación atenta. Finalmente, la inclusión de ítems de codificación opuesta genera más problemas si la intención es interpretarlo como parte de la definición del constructo medido, pero si la intención de estos ítems es la detección de respuestas aquiescentes, entonces estos ítems no deberían entrar en el cálculo de los puntajes de los factores. En ambas situaciones, esto dependerá de los objetivos del usuario para incluirlos.

3.5. Objetivo 3a: Identificar la estructura dimensional del PROPSIT

3.5.1. Introducción

El estudio de la dimensionalidad de una medida tiene un propósito definicional para la teoría y la práctica de uso de cualquier instrumento psicosocial. Sin embargo, existen diferencias conceptuales que deben ser consideradas en la evaluación de la dimensionalidad. Una diferencia entre la *dimensionalidad clásica o estricta* y la *dimensionalidad esencial* ha sido reconocida (Nandakumar, 1991, 2016; Slocum-Gori, & Zumbo, 2011). Esta diferenciación es importante no solo para profundizar la teorización de un constructo, sino para aplicar los presupuestos propios de los procedimientos matemáticas para evaluarlo.

Uno de esos procedimientos para identificar la dimensionalidad es el enfoque no paramétrico, que han demostrado ser eficiente y preciso (Meijer, Tendeiro, & Wanders, 2015), aun en condiciones donde el tamaño muestral relativamente pequeño, presenta habituales desafíos (Watson, Egberink, Kirke, Tendeiro, & Doyle, 2018). Estas ventajas también incluyen la efectividad en escalas con pocos ítems (Crutzen, & Peters, 2017; Dima, 2018; Watson, Egberink, Kirke, Tendeiro & Doyle, 2018). Una de estas metodologías, basadas en la covarianza condicional entre los ítems, es el algoritmo DETECT (*Dimensionality Evaluation to Enumerate Contributing Traits index*; Zhang, & Stout, 1999a), que ha demostrado ser efectivo para identificar correctamente la dimensionalidad (Meijer, Tendeiro, & Wanders, 2015; Van Abswoude, Van der Ark, & Sijtsma, 2004).

Sin embargo, aunque son eficientes en muestras moderadamente pequeñas, otras estrategias de validación de sus resultados son requeridas; por ejemplo, realizar inter-validación dentro de la propia muestra (Monahan, Stump, Finch, & Hambleton, 2007), y la implementación de procedimientos

paramétricos (Bonifay, Reise, Scheines, & Meijer, 2015), que sirven como herramientas de corroboración.

El objetivo del presente estudio fue avanzar en la evaluación de la dimensionalidad del *Instrumento para la Evaluación de Procesos Psicosociales del Trabajo* (PROPSIT; Juárez-García, & Flores-Jiménez, 2020), una medida creada para diferenciar los factores psicosociales del contexto laboral, moderadores activos, efectos en la salud y bienestar, y otras variables psicosociales intervinientes en trabajadores mexicanos. El modelo-marco de esta medida es demandas-recursos laborales (Bakker & Demerouti, 2007; Demerouti, Bakker, Nachreiner, & Schaufeli, 2001; Bakker & Demerouti, 2014), y directamente del modelo en que los factores de riesgo psicosocial y los factores o recursos salutogénicos interactúan con la variabilidad individual del trabajador (Juárez & Flores, 2020), para producir efectos en la salud y bienestar. El PROPSIT fue construido en México, pero es potencialmente generalizable a otros contextos latinoamericanos, por el análisis teórico de los factores con valor ético (i.e., factores consistentemente identificados en varios estudios internacionales (Broetje, Jenny, & Bauer, 2020)).

En la exploración de la dimensionalidad toma en cuenta, entre otras cosas como las presunciones estadísticas de los datos, el tamaño de la muestra (Brown, 2015); por ejemplo, cuando se hace una evaluación paramétrica de la dimensionalidad por medio del modelamiento SEM, se esperan muestras grandes particularmente en medidas multidimensionales. Desde el contexto de una medida multidimensional, sin embargo, se pueden buscar estructuras unidimensionales (Hemker, Sijtsma, & Molenaar, 1995; Sijtsma, Meijer, van der Ark, & Mokken, 2011), esto es, que la evaluación puede hacerse diferenciadamente para cada contenido, en el contexto de otros contenidos. Esto alivia la preocupación del investigador de analizar muestras pequeñas en medidas de poco número de ítems, y que además exploratoriamente es una razonable aproximación a la dimensionalidad que necesariamente no pierde precisión.

Se establecieron hipótesis sobre la dimensionalidad del PROPSIT:

Hipótesis 1. El PROPSIT tendrá una estructura multidimensional con dimensiones oblicuas. De acuerdo al modelo teórico y resultados empíricos de los factores psicosociales del trabajo de Juárez y Flores (2020), existen varias dimensiones no redundantes, independientes y asociadas que caracterizan a los factores de riesgo psicosociales y a los recursos favorables. Por lo tanto, se espera que los ítems representen la unidimensionalidad en su propio puntaje.

Hipótesis 2. Las dimensiones serán estadísticamente diferenciadas. Aunque realísticamente se espera que las dimensiones latentes psicosociales mantengan algún grado de covariabilidad, el grado en que esta asociación varía son consecuencias del error de medición y de la teoría sustantiva, entre otras cosas (Brown, 2015). De acuerdo al estudio de Juárez-García y Flores (2020), las dimensiones latentes en las áreas de factores de riesgo psicosocial (*demandas psicológicas, exigencias físicas y acoso psicológico*) covariaron en $r_{media} = .329$, con correlaciones aparentemente homogéneas. Esto indica que las dimensiones pueden expresarse como variables latentes moderadamente independientes. Por otro lado, por construcción teóricamente guiada (Juárez-García & Flores, 2020) las dimensiones no deberían hallarse altamente relacionadas.

El desarrollo de este objetivo fue evaluado con dos enfoques: no paramétrico y paramétrico. El primero fue exploratorio para obtener la primera aproximación de la dimensionalidad en una muestra relativamente pequeña. El otro enfoque, paramétrico, fue aplicado a una muestra con mayor tamaño.

Método

Participantes

Evaluación no paramétrica. Los participantes fueron muestreados no probabilísticamente; los criterios de elegibilidad fueron mayoría de edad, ser peruano, trabajador con contrato formal, actividad laboral actual, y asentir voluntariamente para participar. Este grupo de participantes fue el grupo de

trabajadores de diferentes carreras y puestos de trabajo formales, analizado en el Objetivo 1c (validez de los ítems del PROPSIT con criterios externos) y cuya distribución aparece en su Tabla 16, y Objetivo 2 (Identificación de sesgos en las respuestas al PROPSIT - FPST).

Evaluación paramétrica. Las muestras de participantes incluidas en este análisis fueron dos: a) los mismos participantes del Objetivo 2 (Identificación de sesgos en las respuestas al PROPSIT - FPST), que consistieron en el grupo de trabajadores de variadas carreras y puestos de trabajo (analizado en la evaluación no paramétrica y descrito en el párrafo anterior), y b) en los docentes de educación básica regular. Sin embargo, las muestras efectivas fueron los participantes que no fueron detectados por posibles respuestas con esfuerzo insuficiente (ver Tabla 21 del Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST).

Instrumentos

Escala de Procesos Psicosociales del Trabajo (Juárez-García, & Flores-Jiménez, 2020).

Medida de factores, efectos y procesos intervinientes del trabajo, para evaluar las condiciones psicosociales laborales y los efectos sobre el bienestar del trabajador. Contiene cuatro secciones generales: Factores Psicosociales del Trabajo (FPST), Efectos psicosociales del trabajo, Proceso Salud-Enfermedad y Otros Factores Psicosociales extra-laborales. El presente estudio validó la sección de FPST, consistente de dos grandes áreas: *Recursos Favorables* (dimensiones de recompensas y desarrollo laboral, control laboral, recursos para hacer el trabajo, clima y apoyo social laboral y congruencia de valores), y *Riesgos Psicosociales*, que consiste en varias dimensiones como Demandas Laborales (subdimensiones: carga y ritmo de trabajo, alta responsabilidad y peligrosidad, jornadas, turnos u horarios, cognitivas o atencionales, emocionales por trato con personas, esfuerzo físico y ambiente físico), Acoso Laboral psicológico y Liderazgo Estresante. En el estudio de los autores

(Juárez-García, & Flores-Jiménez, 2020), las confiabilidades α y ω variaron respectivamente entre .73 y .85, y .70 y .87, con poca distancia entre ellas.

Procedimiento

Recolección de datos. La convocatoria a los participantes se hizo mediante las redes sociales del investigador principal, y dos colaboradoras peruanas, egresadas de la carrera de psicología; las redes sociales usadas fueron WhatsApp y Facebook. La elegibilidad de la selección de los contactos se estableció en la edad (18 años mínimo) y en la actividad laboral (estar trabajando en el momento de la recolección de datos). El primer contacto con los participantes fue para explicar brevemente por escrito, en qué consistió el estudio y el contenido de la encuesta; luego, los participantes recibieron el link por el mismo medio de comunicación. Los participantes fueron también invitados a reenviar el link a sus propios contactos. Los interesados en participar debieron hacer clic en el link, y éste los condujo a tener acceso a la encuesta en la plataforma web. El material fue ordenado en una secuencia única para todos: formulario de consentimiento de participación (con información sobre el objetivo del estudio, la voluntariedad de la participación, la anonimidad de respuesta, la libertad para dejar de responder en cualquier momento, el tratamiento confidencial de los datos, y la ausencia de rastreabilidad del participante). La elegibilidad de los participantes luego de recolectar los datos se orientó a) la pertenencia contractual a una institución laboral peruana en el último empleo, y b) el tiempo de trabajo de mínimo de seis meses.

Análisis. En un marco completamente exploratorio, y debido a la alta y frecuente inconsistencia de los métodos para identificar la dimensionalidad esencial o estricta de medidas psicológicas (Hoyle, & Duvall, 2004; Pearson, Mundfrom, & Piccone, 2013; Slocum-Gori, & Zumbo, 2011), primero se evaluó el número de dimensiones latentes para el contenido del PROPSIT, mediante un *enfoque de consenso o acuerdo* (Knekta, Runyon, & Eddy, 2019; Lüdecke, Ben-Shachar, Patil, & Makowski,

2020) aplicado independientemente a las dos áreas de contenido, factores de riesgo psicosocial y recursos favorables (FPS y FRF, respectivamente). Mediante este método, el número de probables dimensiones ajustado a los datos se estima por varias familias de técnicas (e.g., basadas en la prueba de máxima verosimilitud, en los autovalores, múltiple regresión, parsimonia de información, prueba de Barlett, etc.) y se obtiene el número de dimensiones basado en el consenso de los métodos aplicados. Se aplicó el paquete R *parameters* (Lüdtke, et al., 2020), que extrae los métodos de los paquetes *nFactors* (Raiche, & Magis, 2020) y *psych* (Revelle, 2020). Información detallada sobre los métodos pueden hallarse en las fuentes correspondientes.

Marco no paramétrico. En segundo lugar, cada grupo teórico de ítems fue evaluado sobre su *unidimensionalidad esencial* (Stout, 1987), esto es, la existencia de un factor dominante en presencia de factores secundarios y triviales. Esto se hizo mediante dos enfoques no paramétricos de teoría de respuesta al ítem. El primer enfoque no paramétrico fue el algoritmo DETECT (*Dimensionality Evaluation to Enumerate Contributing Traits index*; Zhang, & Stout, 1999a), aplicado a ítems ordinales (*Poly-DETECT*; Zhang, 2007; Zhang & Stout, 1999). Está basado en la covarianza condicionada entre pares de ítems para detectar si un grupo de ítems es esencialmente unidimensional o multidimensional, pero configurada como una simple estructura. Debido al número pequeño de ítems en cada grupo de contenido, y para no forzar la identificación de contenidos falsamente unidimensionales o multidimensionales (Zhang, 2007, p. 27), se probó la multidimensionalidad de dos contenidos presentados contiguamente en el PROPSIT. Por ejemplo, el contenido *carga y ritmo de trabajo* junto con *alta responsabilidad y peligrosidad*, fueron incluidos conjuntamente en la evaluación de la multidimensionalidad; luego *jornada, turno y horario* junto con *cognitivo – atencionales*, y así para los siguientes contenidos. Debido a la independencia teórica estos contenidos evaluados conjuntamente, se esperó que este conjunto sea detectado como multidimensional y aproximadamente de estructura simple, esto es, que los ítems miden una dimensión dentro de una medida multidimensional.

Las guías interpretativas para DETECT varían en magnitud; por ejemplo, Kim (1994) propuso valores de $< .10$ (unidimensionalidad), $< .5$ (débil dimensionalidad), < 1.0 (moderada dimensionalidad) y > 1.0 (fuerte dimensionalidad); Roussos y Ozbek (2006) propusieron: $< .20$ (unidimensionalidad esencial), $> .20$ (multidimensionalidad débil a moderada), $> .40$ (multidimensionalidad moderada a fuerte), > 1.0 (multidimensionalidad fuerte). En el presente estudio, ambos criterios fueron aplicados. Junto con poly-DETECT, ASSI (*Approximate Simple Structure Index*; Zhang, & Stout, 1999b) y el índice *RATIO* (Zhang, & Stout, 1999b) se aplican complementariamente para detectar estructuras aproximadamente simples, y que indiquen la óptima clasificación de ítems en dimensiones claramente diferenciadas. ASSI y *RATIO* varían entre 0 a 1, y se utilizaron los puntos de corte sugeridos para indicar unidimensionalidad esencial (ASSI $< .25$, *RATIO* $< .36$), desviación la unidimensionalidad (ASSI $> .25$, *RATIO* $> .36$) y evidente simple estructura multidimensional (ASSI y *RATIO* > 1.0). Debido el tamaño de la muestra relativamente pequeña, se implementó una *inter-validación* mediante la segmentación de la muestra en una de entrenamiento y otra de validación (Monahan et al., 2007). La detección no paramétrica de la multidimensionalidad en el PROPSIT se completó con un procedimiento visual de búsqueda de la estructura, mediante un análisis de cluster jerárquico (método Ward), y representado visualmente con dendogramas. Se usó el programa *sirt* (Robitzsch, 2019).

Debido que el uso conjunto otros procedimientos para verificar los resultados de DETECT ha probado ser una estrategia efectiva y segura (Bonifay, et al., 2015; van Abswoude, van der Ark, & Sijtsma, 2004), sin salir de un enfoque no paramétrico, en segundo lugar se implementó *Automated Item Selection Procedure* (AISP; Mokken, 1971), un procedimiento que segmenta un conjunto de ítems en subgrupos con propiedades del *modelo de homogeneidad monotónica*, y lograr así una escala *Mokken* para el puntaje (Van der Ark, 2012). El procedimiento inicia buscando ítems con coeficientes de escalabilidad H mínimos de .30 (Sijtsma K, van der Ark LA., 2017), y continua hasta otros niveles H establecidos a priori. En un enfoque exploratorio para buscar estructuras o subescalas, AISP conduce

a identificar un grupo de ítems escalables con un mínimo nivel satisfactorio, es decir, una escala donde la variable latente puede ser psicométricamente isomórfica con el puntaje observable (Mokken, 1971; Sijtsma, & van der Ark, 2017). Los resultados de AISP definen una escala Mokken pero no necesariamente la propiedad de unidimensionalidad (Smits, Timmerman, & Meijer, 2012), por lo tanto, estos resultados fueron contrastados y evaluados conjuntamente con el algoritmo poly-DETECT. Se utilizó el programa R *mokken* (van der Ark, 2012).

Marco paramétrico. Debido que existe una estructura predefinida de configuración de los ítems del PROPSIT, el estudio de las dimensiones de los ítems del PROPSIT se efectuó con el procedimiento de *análisis factorial semi-confirmatorio* (Browne, 1972; McDonald, 2005); este efectúa una rotación de tipo Procusto hacia una matriz a priori definida (*matriz target*) sobre la relación de los ítems en los factores del modelo de medición. Este procedimiento fue descrito en el Objetivo 3 (análisis de sesgos de respuesta en el PROPSIT - FPST), con las mismas especificaciones, estimadores y decisiones de ajuste. En los presentes resultados, se identificó la estructura interna luego de remover los datos con posibles respuestas con esfuerzo insuficiente, los resultados se orientan a describir con detalle los parámetros de ajuste (coeficientes de congruencia), las cargas factoriales, las correlaciones factoriales. Se usó el mismo programa Factor (Ferrando & Lorenzo-Seva, 2018).

3.5.2. Resultados

Evaluación no paramétrica

Número de dimensiones. Para la aplicación de este algoritmo, únicamente los métodos que lograron resultados sin problemas de estimación fueron reportados. De este modo, para los ítems de Riesgos Psicosociales, 1 dimensión fue respaldada por 4 métodos de un total de 20 (20.0%; prueba t – regresión múltiple, valor p – regresión múltiple, TLI, RMSEA). Sin embargo, 3 métodos (15.0%) apoyaron elegir 3 factores (Cattell-Nelson-Gorsuch, error estándar de autovalores, y EGA - TMFG), así

como 4 factores (métodos regresión múltiple, Exploratory Graph Analysis – Glasso y BIC – ajuste) y 5 factores (métodos: coordenadas óptimas, análisis paralelo y regla de Kaiser).

Para los ítems de Recursos Favorables, de los 25 métodos que lograron ser estimados para la identificación del número de factores aplicados, 6 (24%; Cattell-Nelson-Gorsuch, coordenadas óptimas, análisis paralelo, regla de Kaiser, EGA – TMFG y Very Simple Structure – complejidad 2) indicaron 3 dimensiones latentes, y otros 6 (24%; regresión múltiple, error estándar de autovalores, EGA – glasso, MAP de Velicer, BIC - parsimonia de información, y BIC - prueba de ajuste) indicaron 4 dimensiones latentes.

Algoritmo poly-DETECT. En la Tabla 26 se muestran los coeficientes sumarios del algoritmo poly-DETECT y coeficientes complementarios (ASSI y RATIO). Se observa que cada conjunto es una agrupación multidimensional de ítems en los Factores de Riesgo Psicosocial ($M_{\text{poly-DETECT}} = 26.08$) y más grande que los ítems de Recursos Favorables, pero en conjunto, todos los poly-DETECT fueron grandemente > 1.0 . La multidimensionalidad en cada conjunto corresponde con las expectativas de dimensionalidad declaradas en el estudio. Respecto a ASSI y RATIO, estructuras multidimensionales simples fueron detectadas fuertemente en el conjunto 3, 4, y moderadamente en los conjuntos 1 y 2; por otro lado, el conjunto 5 estuvo debajo del punto de corte en ambos indicadores, pero ligeramente en ASSI; el conjunto 5 identificó marginalmente unidimensionalidad en ambas áreas de contenido (Recompensas y desarrollo laboral + Tareas benefician a personas y sociedad), sugiriendo elevada covariación entre ambos.

La distribución de ítems de FRP en las subescalas se muestra en la Tabla 27, derivado poly-DETECT y de MSA. En el conjunto 2 (*Jornada, turno y horario + Cognitivas/atencionales*), y conjunto 3 (*Demandas emocionales + esfuerzo/ambiente físico*) los ítems se distribuyeron de acuerdo a su estructura teórica, indicando convergencia con sus valores ASSI y RATIO (= 1). Por otro lado, el

conjunto 1 fue moderadamente eficaz debido que el ítem 2 no se mantuvo en *carga y ritmo de trabajo*; y el conjunto 4 unificó todos los ítems en una sola dimensión, excepto el ítem 22 cuya codificación se invierte. Por otro lado, en Recursos Favorables (FRF), casi completamente los ítems mostraron consistencia teórica en su estructura esperada (Tabla 28): en el conjunto 1, los ítems recodificables formaron un cluster separado de su escala esperada; y en el conjunto 2, los ítems se agruparon completamente también teóricamente. En la Figura 15 se presentan los dendogramas de clasificación de los ítems de FRP y RFR.

Tabla 26

Resultados no paramétricos (algoritmo DETECT) para la dimensionalidad

Conjuntos evaluados	Poly-DETECT	ASSI	RATIO
Factores de Riesgo Psicosocial			
Conjunto 1: Carga y ritmo de trabajo + Alta responsabilidad y peligrosidad	40.91	.6	.88
Conjunto 2: Jornada, turno y horario + Cognitivas / atencionales	15.40	.46	.71
Conjunto 3: Emocionales + Esfuerzo físico + Ambiente físico	46.41	1	1
Conjunto 4: Acoso laboral + Liderazgo estresante	16.43	1	1
Factores de Recursos Favorables			
Conjunto 5: Recompensas y desarrollo laboral + Tareas benefician a personas y sociedad	4.51	.24	.24
Conjunto 6: Recursos para realizar el trabajo + Congruencia de valores	5.28	.52	.65

Nota. ASSI: Approximate Simple Structure Index.

Tabla 27

Resultados no paramétricos de la distribución de los ítems de Riesgos Psicosociales (FRPS)

	Clusters	AISP (MSA)		
		.3	.4	.5
CONJUNTO 1				
RP: Demandas laborales				
Carga y ritmo de trabajo				
1. Carga de trabajo	1	1	1	1
2. Trabajar rápido	1	1	1	1
3. Interrupciones para tareas a tiempo	2	2	2	0
Alta responsabilidad y peligrosidad				
4. Tomar acciones importantes	2	2	2	2
5. Peligros	2	2	2	2
CONJUNTO 2				
Jornada, turno y horario				
6. Horas extras, horarios prolongados	1	1	1	0
7. Rotación/cambio de turno	1	2	2	2
8. Debo laborar jornadas en turno nocturno	1	2	2	2
Cognitivas / atencionales				
9. Atención prolongada en tareas	2	1	1	1
10. Atender dos o más tareas	2	1	1	1
11. Esfuerzo mental	2	1	1	1
CONJUNTO 3				
Emocionales				
12. Emociones negativas de otras personas	1	2	2	2
13. Trato con personas desgastante	1	2	2	2
14. Mostrar emociones diferentes a las que siento	1	2	2	2
Esfuerzo físico				
15. Mucho esfuerzo físico	2	1	1	1
16. Posiciones incómodas	2	1	1	1
Ambiente físico				
17. Condiciones ambientales adversas	2	1	1	1
CONJUNTO 4				
RP: acoso laboral				
Por superiores				
18. Maltrato psicológico (superiores)	1	1	1	1
Por compañeros				
19. Maltrato psicológico (compañeros)	1	1	1	1
RP: Liderazgo estresante				
Supervisión controlante				
20. Demasiado control	1	1	1	1
21. Reglas y normas exageradas	1	1	1	1
Retroalimentación negativa (supervisor)				
22. Retroalimentación adecuada de supervisores (R)	2	0	0	0

Nota. AISP: Automated Item Selection Procedure. MSA: Mokken Scaling Analysis. Cluster: clasificación derivada del análisis de clusters (método Ward).

Tabla 28

Resultados no paramétricos de la distribución de los ítems de Factores de Recursos Favorables (FRF)

	Cluster	AISP (MSA)		
		.3	.4	.5
CONJUNTO 1				
RF: Recompensas y desarrollo laboral				
Justicia organizacional				
23. Trabajo justo y equitativa.	1	1	1	1
Salario motivante				
24. Salario motivante	1	1	1	1
Reconocimiento por el trabajo				
25. Trabajo apreciado y reconocido	1	1	1	1
Posibilidades de desarrollo laboral/profesional (estatus)				
26. No oportunidades de crecer profesionalmente (R)	2	2	2	2
Seguridad en la conservación del empleo				
27. Pienso que puedo perder este empleo próximamente (R)	2	2	2	2
Tarea gratificante y con trascendencia				
28. Trabajo agradable y gratificante	1	1	1	1
29. Tareas benefician a personas y sociedad	3	1	1	1
RF: Control laboral y contenido de la tarea				
Influencia – autonomía en el trabajo				
30. Libre en decidir el trabajo	3	1	1	1
Utilización de habilidades en el trabajo				
31. Utilizo habilidades	3	1	1	1
32. Desarrollo de capacidades	3	1	1	1
Trabajo variado (no monótono)				
33. Actividades muy variadas	3	1	1	1
Funciones y roles claros				
34. Funciones y tareas muy claras	3	1	1	1
CONJUNTO 2				
RF: Recursos para realizar el trabajo				
Recursos materiales, equipos, herramientas para el trabajo				
35. Materiales necesarios y adecuados	1	1	1	1
Capacitación, adiestramiento				
36. Capacitaciones necesarias	1	1	1	1
RF: Clima y apoyo social laboral				
Apoyo de compañeros				
37. Compañeros apoyan	2	1	1	1
Apoyo de jefes				
38. Superiores apoyan	2	1	1	1
Clima de unión				
39. Clima de unión/colaboración	2	1	1	1
RF: congruencia de valores				
Congruencia individuo - organización				
40. Coincidencias mis valores y los de la organización	3	1	1	1
Congruencia individuo – compañeros – superiores				
41. Coincidencia mis valores – compañeros	3	1	1	1

Nota. AISP: Automated Item Selection Procedure. MSA: Mokken Scaling Analysis. Cluster: clasificación derivada del análisis de clusters (método Ward).

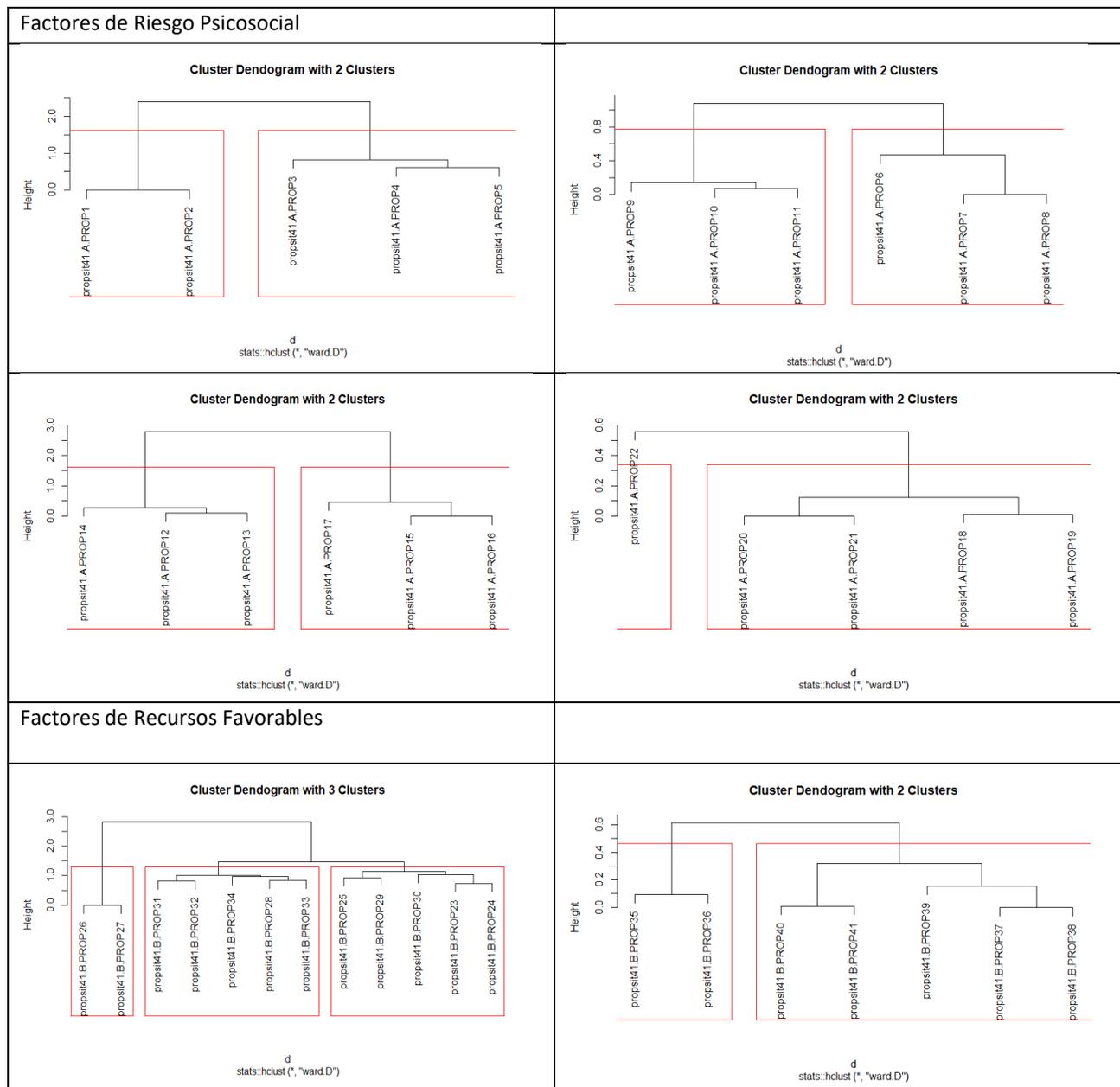


Figura 15

Dendrograma del análisis de clusters de los ítems del PROPSIT

Automated Item Selection Procedure (AISP). El algoritmo AISP para el contenido de Riesgos Psicosociales (Tabla 27) en general fue concordante con los resultados del procedimiento poly-DETECT. Concretamente, el conjunto 3 fue completamente coincidente con el modelo teórico, mientras que el conjunto 4 fue concordante con los resultados de Juárez-García y Flores (2020),

excepto el ítem que requiere recodificación. Y los conjuntos 1 y 2 fueron distinguidos apropiadamente, excepto los ítems 3 (*Carga y ritmo de trabajo*) y 6 (*Jornada, turno y horario*). Las discrepancias halladas en estos ítems fueron resaltadas en el nivel $H > .50$, un nivel de escalabilidad considerado como alto.

Respecto a los Recursos Favorables (Tabla 28), se observó discrepancia frente a los resultados de poly-DETECT, debido que AISP fue más propenso a detectar escalas fusionadas en una sola, en cada conjunto evaluado. Los ítems que requieren recodificación (ítems 26 y 27) fueron diferenciados del resto de ítems, tal como ocurrió también en los resultados poly-DETECT.

Evaluación paramétrica

Factores de Riesgo Psicosocial. Se hicieron dos iteraciones, la primera describe la evaluación de la estructura teórica-empírica del PROSIT, y la segunda expone la afinación del modelo.

Primera iteración. En la Tabla 29 se muestran los resultados de la factorización semi-confirmatoria, cuya matriz target está presentada en la Tabla 22 del Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST. En la columna etiquetada *C* se observa la congruencia entre la configuración hipotética de cada ítem, y la configuración resultante. Considerando esta medida de ajuste, la configuración en la muestra de educadores fue altamente inconsistente, excepto el factor F7 que mostró mejor ajuste; en la muestra de trabajadores en general se halló mayor consistencia, excepto en los factores F1, F3 y F4. En la muestra total, la congruencia estuvo debajo del criterio de aceptabilidad ($< .85$), y los factores F5, F6 y F7 mostraron mayor congruencia.

El contenido de acoso laboral y liderazgo estresante ahora están contenidos en un factor, excepto el ítem 22. Este ítem 22 se debe recodificar para ser interpretado junto con el resto de los ítems de su grupo inicial (ie, liderazgo estresante), pero en los resultados se moviliza hacia otras posibles

dimensiones con mayor carga comparada con su carga factorial en su grupo inicial, y en cada muestra, esta movilidad no es constante.

Los grupos de ítems concernientes a esfuerzo físico y ambiente físico, se solaparon en un solo factor, tal como ocurrió también en el estudio factorial de Juárez-García y Flores (2020), vinculando las demandas físicas exigidas al trabajador, y las características físicas del ambiente del trabajo. Los ítems que describen respuestas emocionales y situaciones interpersonales se mantuvieron independientes del resto, y por lo tanto permanecieron diferenciadas. El contenido de la demanda cognitiva/atencional fue reunido con el contenido carga y ritmo de trabajo, representando una dimensión sobre la respuesta cognitiva frente a demandas laborales de rapidez y monto de trabajo asignado.

Los ítems que conceptualmente fueron elaborados para medir la peligrosidad e importancia, así como rotación y largas horas de trabajo nocturno, fueron inestables en permanecer integrado, y se movilizaron de manera inconsistente entre el resto de los contenidos. Estos ítems, y algunos otros, también mostraron un monto de complejidad factorial en por lo menos dos dimensiones, indicando que los ítems pueden representar más de una dimensión. Por ejemplo, el ítem 3 (Interrupciones para tareas a tiempo), e ítem 5 (actividades que implican peligro); estos fueron señalados con un subrayado cuando fueron iguales o mayores a .30. Otros ítems también mostraron complejidad cuando sus cargas factoriales en otros factores fueron alrededor de .20 (por ejemplo, ítems 18, 19 y 20), pero en un grado menor.

En conclusión, se detectaron grupos de ítems teóricamente coherentes, moderadamente estables en la muestra total y en las dos submuestras separadas, pero también ocurrieron ítems con complejidad factorial y movilidad dimensional, no replicables en las dos muestras.

Segunda iteración. Para la siguiente iteración de evaluación de la estructura interna, se tomó en cuenta a) el objetivo del presente proyecto (adaptar el PROPSIT a un contexto diferente con un grado

de generalización de sus constructos), y b) los resultados obtenidos de la dimensionalidad del PROPSIT (estabilidad de los ítems en las dimensiones, y grado de complejidad factorial); por lo tanto, se propuso un modelo con los ítems más estables y menor complejidad factorial. Los ítems 3, 4, 5, 7, 8, 9 y 22 fueron removidos previamente debido a: a) la inconsistencia empírica mostrada en la primera iteración, y b) que estos ítems convergieron con los ítems removidos en el análisis final de Juárez-García y Flores (2020).

La configuración factorial en esta iteración se muestra en la Tabla 30, con el encabezado *Target revisada*. El ajuste de este modelo modificado en la muestra total se observa en el encabezado *Estimación y Ajuste* de la Tabla 30, en se muestra una estructura con altas cargas factoriales en las propias dimensiones ($> .53$) y predominantemente cargas factoriales bajas ($< .15$) en las dimensiones divergentes. Todos los coeficientes de congruencia para los ítems, dimensiones y solución factorial total, mostraron ser mayores a .93. De acuerdo al contenido de los ítems, y parecido con los factores definidos en el estudio de Juárez-García y Flores-Jiménez, los factores retenidos en esta solución final pueden ser identificados como *demandas psicológicas* (F1), *exigencias físicas* (F2), *acoso psicológico* (F3), y *demandas emocionales* (F4). Las correlaciones entre estas dimensiones muestran tendencia hacia las relaciones moderadas y con suficiente independencia entre ellas.

Factores Recursos Favorables. Como en los resultados de la estructura interna de los factores de riesgo psicosocial (FRPS), para los factores de recursos favorables se hicieron dos iteraciones: la primera describe la evaluación de la estructura teórica-empírica del PROPSIT, y la segunda expone la afinación del modelo.

Primera iteración. La matriz target para esta iteración se encuentra en la Tabla 23 del Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST, y los resultados obtenidos se muestran en la Tabla 31. En la muestra total, más de la mitad de los ítems obtuvieron ajuste apropiado, medido por los

coeficientes de congruencia ($C > .90$), y los factores F2 y F4 mostraron el mejor ajuste comparado con los factores F1 y F3. El ajuste total no alcanzó un nivel aceptable ($C < .85$), y esto sugiere una reestructuración con base empírica del modelo de medición. Este patrón fue parcialmente reproducido en las muestras por separado, donde en ambas muestras el ajuste fue excesivamente deteriorado en los factores F1 (muestra de educadores) y F3 (muestra en general).

La correlación Pearson entre los coeficientes C de los ítems en cada muestra fue .536, indicando un moderado parecido. La aparente fuente de discrepancia se encontró en los ítems 23, 24, 25 y 35, en que se observa alta complejidad factorial e inestabilidad en su configuración, así como los ítems 28 y 29 en las submuestras (trabajadores en general y educadores); los ítems 35 y 36 también mostraron complejidad factorial y movilidad hacia otras dimensiones, especialmente hacia la dimensión de recompensas, como ocurrió en el estudio de Juárez-García y Flores-Jiménez.

Segunda iteración. Para la siguiente iteración de evaluación de la estructura interna, se tomó en cuenta a) el objetivo del presente proyecto (adaptar el PROPSIT a un contexto diferente con un grado de generalización de sus constructos), y b) los resultados obtenidos de la dimensionalidad del PROPSIT (estabilidad de los ítems en las dimensiones, y grado de complejidad factorial). Por lo tanto, se propuso un modelo modificado para los Factores de Recursos Favorables, que aparece en la Tabla 32, bajo el encabezado Matriz Target revisada. Los ítems 26 y 27 fueron removidos debido a) al fraseo contrario al constructo, que pueden inducir por varianza de método, y b) respectivamente ambos ítems fueron removidos por mostrar baja carga factorial en el estudio de Juárez-García y Flores-Jiménez (2020). También fueron removidos los ítems 28 y 29 por: a) bajas cargas factoriales y complejidad factorial, y b) también removidos en el estudio de Juárez-García y Flores-Jiménez (2020). Todos los coeficientes de configuración, en el nivel de los ítems, factores y solución total, fueron de magnitud alta ($C \geq .95$), sugiriendo que la configuración revisada de RF identificó los factores óptimamente. La simplicidad factorial de los ítems también es una fuente de esta alta congruencia. El contenido de estos

factores se empareja con el estudio de Juárez-García y Flores-Jiménez (2020), y son identificados como *recompensas y recursos* (F1), *clima de apoyo social* (F2), y *control laboral* (F3).

Se concluye que, similarmente al estudio de Juárez-García y Flores-Jiménez (2020), los factores retenidos en los riesgos psicosociales representan los constructos de demandas psicológicas, exigencias físicas, acoso psicológico, y adicionalmente un factor que emergió independiente del resto fue demandas emocionales. Mientras que en recursos favorables emergieron los factores de recompensas y recursos, clima de apoyo social, y control laboral. En resumen, la re-ubicación de los ítems en las dimensiones mostró mayor validez estructural, así como una configuración factorial simple.

3.5.3. Discusión

Evaluación no paramétrica

Cuando se evaluó el número de dimensiones latentes, la variabilidad de respuestas al número de dimensiones latentes por los métodos basados en el modelamiento lineal, es impresionante grande, y el consenso ayudó a poner un rango razonable que moderadamente se empata con las dimensiones teóricas. Como respuesta al objetivo del presente estudio, los resultados sobre el número de dimensiones se distancian poco con el número de dimensiones del estudio de Juárez-García y Flores-Jiménez (2020), y la aproximación exploratoria en este punto sugirió que una aparente replicabilidad. Con resultados exploratorios como éstos, una decisión que balancea la evidencia empírica y la racionalidad teórica es extraer el número de dimensiones identificado ± 1 (Hoyle, & Duvall, 2004). Pero con el marco semi-confirmatorio utilizado, esta opción es aplicable a la posible reestructuración de modelo de medición del PROPSIT, como se describe más adelante.

Tabla 29

Factores de Riesgo Psicosocial: Análisis factorial, primera iteración en la muestra total y submuestras (muestra limpia)

	Educadores (n = 121)								General (n = 184)								Total (n = 305)							
	F1	F2	F3	F4	F5	F6	F7	C	F1	F2	F3	F4	F5	F6	F7	C	F1	F2	F3	F4	F5	F6	F7	C
1. Carga de trabajo	.31	-.17	.18	<u>.32</u>	.07	-.15	.08	.57	.14	.01	-.11	.81	-.08	.10	.08	.16	<u>.35</u>	.05	-.01	.70	-.15	.00	.04	.43
2. Trabajar rápido	1.03	.01	.15	-.04	.04	-.26	.01	.96	.15	-.08	-.08	.78	-.06	.17	.08	.18	.44	.09	-.03	.69	-.15	-.01	.06	.52
3. Interrupciones para tareas a tiempo	.51	-.17	-.27	.10	-.04	.42	.04	-.23	.00	.28	-.25	-.13	.38	.16	.18	.47	.71	.69	-.13	-.27	.12	.00	.04	.66
4. Tomar acciones importantes	-.01	<u>.32</u>	.20	-.11	<u>.31</u>	.17	-.10	.60	-.15	.76	.14	.23	.14	.04	-.19	.89	-.02	<u>.32</u>	<u>.31</u>	-.05	.24	.14	-.21	.55
5. Peligros	.00	-.01	.56	-.06	-.07	<u>.37</u>	-.14	-.01	.07	.89	.07	-.01	-.16	.04	.02	.97	.09	.68	<u>.37</u>	.04	-.06	.04	-.11	.86
6. Horas extras, horarios prolongados	.13	.00	.43	.03	-.22	.52	-.06	.59	.12	-.08	.15	.41	.03	.16	.08	.30	.10	.09	.16	.50	-.11	.13	.08	.27
7. Rotación/cambio de turno	.14	.42	<u>.35</u>	.06	.00	-.13	.08	.59	.03	.04	.67	-.01	.10	-.03	.04	.98	-.10	.17	.46	.14	.06	-.08	.09	.85
8. Jornadas en turno nocturno	-.11	-.07	.80	.11	-.18	-.20	.20	.90	-.04	.13	.67	-.11	.02	-.02	.07	.96	-.12	.29	.53	.09	-.09	-.12	.13	.81
9. Atención prolongada en tareas	-.11	.00	.26	.43	.04	.09	-.11	.80	.14	.04	.12	.76	-.05	.00	-.04	.96	-.08	-.07	.16	.78	.01	.02	-.06	.96
10. Atender dos o más tareas	.03	.03	-.06	1.04	-.10	.04	-.01	.99	.07	-.01	.01	.83	.07	.10	-.05	.98	.15	-.03	.09	.79	.00	.03	-.06	.97
11. Esfuerzo mental	.08	-.29	.26	<u>.25</u>	.48	-.25	-.03	.34	.00	.12	-.15	.82	.05	-.31	.07	.91	.05	-.12	-.02	.63	.23	-.17	-.02	.89
12. Emociones negativas otras personas	-.01	.13	-.14	.17	<u>.77</u>	-.12	-.02	.93	-.02	.22	-.08	.19	.72	-.09	-.04	.91	.02	.05	.02	.17	.79	-.09	-.07	.96
13. Trato personas desgastantes	-.06	.05	-.13	.08	.48	<u>.37</u>	.21	.72	.06	-.09	.15	.03	.71	-.03	.12	.95	-.07	-.03	.00	.14	.59	.08	.23	.89
14. Mostrar emociones diferentes	.08	-.05	-.09	-.17	<u>.72</u>	.19	-.04	.92	-.06	-.04	.04	-.11	.87	.06	-.04	.98	.07	.13	-.07	-.08	.71	.04	.01	.96
15. Mucho esfuerzo físico	.01	.04	-.02	.11	.41	.35	.02	.64	.00	-.04	-.01	-.09	.08	1.02	-.15	.98	-.03	-.05	-.08	-.04	.11	.92	-.08	.98
16. Posiciones incómodas	-.02	.09	.04	.03	.15	.71	.03	.96	.03	-.02	.04	-.04	-.02	.85	.03	.99	.01	.00	-.02	.06	-.07	.88	.04	.99
17. Condiciones ambientales adversas	-.05	-.10	-.02	-.05	-.04	.74	.06	.98	.00	<u>.38</u>	-.09	-.11	-.23	.44	.25	.63	.02	.25	.02	-.19	-.08	.46	.13	.79
18. Maltrato psicológico (superiores)	.05	.13	.03	-.06	-.11	.03	<u>.87</u>	.97	<u>-.37</u>	-.10	.05	.12	.02	.02	1.03	.93	.14	-.26	.35	-.24	.01	.00	.84	.85
19. Maltrato psicológico (compañeros)	-.08	.16	.04	.00	.02	-.03	<u>.77</u>	.97	-.22	-.02	.04	.08	-.05	.02	.92	.96	-.01	-.24	.24	-.18	.06	.06	.74	.88
20. Demasiado control	-.02	-.20	.06	.00	.17	.08	<u>.69</u>	.92	.62	-.10	.06	-.19	.02	.00	.57	.65	-.10	.14	-.29	.23	.03	.04	.82	.89
21. Reglas y normas exageradas	.07	-.17	.13	-.03	.11	-.01	<u>.68</u>	.94	<u>.36</u>	.17	-.05	-.04	.07	-.10	.52	.77	-.01	.21	-.16	.16	.05	-.05	.69	.91
22. Retroalimentación supervisores (R)	-.11	.19	.10	-.10	<u>.35</u>	-.08	-.14	.29	.02	-.04	.04	.57	-.03	.01	-.14	.22	-.10	-.27	.04	.46	.08	.11	-.13	.22
Ajuste (C)	.77	.1	.71	.79	.77	.7	.89	.71	.22	.84	.81	.7	.92	.9	.86		.57	.78	.62	.68	.92	.93	.89	.78

Nota. En negrita: ítem con fuerte carga factorial en su dimensión, o con satisfactorio nivel de congruencia. En subrayado, ítems con carga en otra dimensión, en $\geq .30$. C: coeficiente de congruencia; en negrita aparecen coeficientes de congruencia que cumplen con el criterio ($C > .85$).

Factores de Riesgo Psicosocial. F1: carga y ritmo de trabajo. F2: responsabilidad y peligrosidad; F3: jornada, turno y horario. F4: cognitivas – atencionales. F5: emocionales. F6: esfuerzo físico. F7: acoso psicológico.

Tabla 30

Factores de Riesgo Psicosocial: modelo final (muestra limpia, n = 305)

	Matriz Target revisada				Estimación y ajuste				
	F1	F2	F3	F4	F1	F2	F3	F4	C
1. Carga de trabajo	9	0	0	0	.86	-.12	-.03	.05	.99
2. Trabajar rápido	9	0	0	0	.91	-.09	-.03	.06	.99
6. Horas extras, horarios prolongados	9	0	0	0	.55	-.11	.23	.12	.89
10. Atender dos o más tareas	9	0	0	0	.86	-.02	.06	-.08	.99
11. Esfuerzo mental	9	0	0	0	.62	.19	-.23	-.05	.90
12. Emociones negativas otras personas	0	9	0	0	.11	.79	-.05	-.07	.98
13. Trato personas desgastantes	0	9	0	0	.08	.60	.08	.19	.94
14. Mostrar emociones diferentes	0	9	0	0	-.11	.77	.08	-.02	.98
15. Mucho esfuerzo físico	0	0	9	0	.00	.13	.82	-.13	.98
16. Posiciones incómodas	0	0	9	0	.11	-.05	.92	-.04	.99
17. Condiciones ambientales adversas	0	0	9	0	-.18	.01	.54	.17	.90
18. Maltrato psicológico (superiores)	0	0	0	9	-.05	-.09	-.02	.93	.99
19. Maltrato psicológico (compañeros)	0	0	0	9	-.09	-.07	-.02	.87	.99
20. Demasiado control	0	0	0	9	.13	.13	.00	.62	.96
21. Reglas y normas exageradas	0	0	0	9	.10	.16	-.03	.61	.96
Ajuste (C)	-	-	-	-	.96	.95	.94	.95	.95
Correlación									
F1	-	-	-	-	1				
F2	-	-	-	-	.32	1			
F3	-	-	-	-	.28	.37	1		
F4	-	-	-	-	.20	.36	.50	1	

Nota. C: coeficiente de congruencia; en negrita aparecen coeficientes de congruencia que cumplen con el criterio ($C > .85$). En negrita: ítem con fuerte carga factorial en su dimensión, o con satisfactorio nivel de congruencia.

Factores de riesgo psicosocial revisados. F1: demandas psicológicas. F2: demandas emocionales. F3: exigencias físicas. F4: acoso psicológico.

Tabla 31

Factores de Recursos Favorables: Análisis factorial, primera iteración en la muestra total y submuestras (muestra limpia, n = 305)

	Total (n = 305)					Educadores (n = 121)					General (n = 184)				
	F1	F2	F3	F4	C	F1	F2	F3	F4	C	F1	F2	F3	F4	C
23. Trabajo justo y equitativa.	-.16	.11	.17	<u>.41</u>	.33	.21	.04	-.12	.52	.37	.59	-.26	<u>.40</u>	.16	.75
24. Salario motivante	-.14	-.01	.86	-.15	.16	.59	-.15	.15	-.01	.94	.91	-.10	<u>.37</u>	-.27	.88
25. Trabajo apreciado y reconocido	-.10	.19	.51	.11	.19	.72	.13	.02	.06	.97	.62	-.08	<u>.39</u>	.03	.84
26. Sin oportunidades profesionales (R)	.77	.19	.09	.01	.96	-.01	-.05	.26	-.15	.03	-.97	.07	.71	.03	.80
27. Perder este empleo (R)	.58	-.13	.16	.03	.93	.04	-.22	<u>.31</u>	-.27	-.08	-1.18	-.11	<u>1.16</u>	-.05	.71
28. Trabajo agradable y gratificante	-.14	.44	-.05	<u>.37</u>	.23	-.03	.21	-.03	<u>.53</u>	-.05	.19	<u>.45</u>	.03	.24	.35
29. Tareas benefician a personas y sociedad	.02	.68	-.20	.22	-.03	-.21	<u>.54</u>	-.12	.27	-.32	.06	<u>.34</u>	.40	.01	.11
30. Libre en decidir el trabajo	.02	.71	.00	.11	.98	.07	.68	.03	-.03	.99	-.11	.66	.12	.22	.92
31. Utilizo habilidades	.03	1.03	-.01	-.08	.99	-.10	.99	.09	-.08	.98	-.22	1.17	.04	-.05	.98
32. Desarrollo de capacidades	-.03	1.01	-.00	-.11	.99	.07	.97	.01	-.09	.99	-.04	1.08	-.01	-.09	.99
33. Actividades muy variadas	-.06	.87	.09	-.15	.97	.00	.83	.05	-.08	.99	.17	.87	-.06	-.10	.97
34. Funciones y tareas muy claras	.05	.96	.07	-.09	.99	-.05	.95	.07	-.15	.98	.16	.71	.20	-.11	.92
35. Materiales necesarios y adecuados	.04	.17	.69	.05	.96	.20	.29	.53	.06	.82	.19	<u>.35</u>	.09	<u>.31</u>	.18
36. Capacitaciones necesarias	-.06	-.01	.64	.17	.95	.02	.14	.62	.28	.89	.62	-.12	.07	<u>.30</u>	.10
37. Compañeros apoyan	.06	-.17	.16	.81	.95	-.27	-.19	.21	.82	.89	-.05	-.10	.15	.90	.97
38. Superiores apoyan	.01	-.17	-.03	1.00	.98	-.16	-.13	.17	.86	.95	-.12	-.16	.03	1.11	.98
39. Clima de unión/colaboración	-.03	-.00	-.00	.85	.99	.10	-.12	-.06	.80	.97	-.07	.05	-.01	.95	.99
40. Coincidencia de valores con organización	-.01	.07	.01	.79	.99	.23	.01	-.09	.64	.93	-.09	.18	-.02	.88	.97
41. Coincidencia de valores con compañeros	-.02	-.10	.06	.82	.98	.17	-.05	-.01	.61	.96	.11	-.16	-.04	.95	.97
Ajuste (C)	.70	.90	.65	.93	.78	.42	.92	.81	.86	.74	0.8	0.9	0.07	0.95	.77

Nota. En negrita: ítem con fuerte carga factorial en su dimensión, o con satisfactorio nivel de congruencia. En subrayado, ítems con carga en otra dimensión, en $\geq .30$. C: coeficiente de congruencia; en negrita aparecen coeficientes de congruencia que cumplen con el criterio ($C > .85$).

Factores Recursos Favorables. F1: recompensas y desarrollo profesional. F2: Control laboral y contenido de la tarea. F3: recursos para realizar el trabajo. F4: clima y apoyo social laboral.

Tabla 32

Factores de Recursos Favorables: modelo final (n = 305)

	Matriz Target revisada			Estimación y ajuste			
	F1	F2	F3	F1	F2	F3	C
24. Salario motivante	9	0	0	1.08	-.10	-.30	.95
25. Trabajo apreciado y reconocido	9	0	0	.66	.15	.00	.97
30. Libre en decidir el trabajo	0	9	0	-.01	.67	.15	.97
31. Utilizo habilidades	0	9	0	-.08	.99	.02	.99
32. Desarrollo de capacidades	0	9	0	-.04	1.00	-.02	.99
33. Actividades muy variadas	0	9	0	.11	.83	-.09	.98
34. Funciones y tareas muy claras	0	9	0	.06	.86	-.02	.99
35. Materiales necesarios y adecuados	9	0	0	.82	.03	-.02	.99
36. Capacitaciones necesarias	9	0	0	.86	-.12	.04	.98
37. Compañeros apoyan	0	0	9	.16	-.15	.75	.95
38. Superiores apoyan	0	0	9	-.06	-.07	.94	.99
39. Clima de unión/colaboración	0	0	9	-.02	.09	.79	.99
40. Coincidencia de valores con organización	0	0	9	-.05	.16	.80	.97
41. Coincidencia de valores con compañeros	0	0	9	.01	-.01	.81	1.00
Ajuste (C)				.97	.97	.97	.97
Correlación							
F1	-	-	-	1			
F2	-	-	-	.69	1		
F3	-	-	-	.80	.68	1	

Nota. C: coeficiente de congruencia; en negrita aparecen coeficientes de congruencia que cumplen con el criterio (C > .85). En negrita: ítem con fuerte carga factorial en su dimensión, o con satisfactorio nivel de congruencia.

Factores de Recursos Favorables: F1: recompensas y recursos. F2: control laboral. F3: clima de apoyo social.

Por otro lado, los resultados no paramétricos indicaron que los contenidos del PROPSIT son suficientemente multidimensionales para describir los factores psicosociales del trabajo diferenciados. Esta conclusión debe ser atemperada con la metodología aplicada, esto es, en conglomerados pequeños de ítems, mayores a 1, y no como como conjunto total de ítems. La manera de analizar los ítems fue necesario para reducir la capitalización del azar en la relativa pequeña muestra analizada.

La diferenciación de las dimensiones convergió moderadamente entre los métodos no paramétricos aplicados, porque DETECT y AISP-Mokken separaron los contenidos específicos en cada conglomerado analizado, y ambos fueron sensibles a la separación de los ítems fraseados negativamente (ítem 22 en FRPS, e ítems 26 y 27). Debido a esta discrepancia entre ambos métodos, los siguientes pasos en la definición de la dimensionalidad del PROPSIT en la muestra peruana requiere. En este punto, se encontró que los ítems que requieren ser recodificados para obtener un puntaje con el resto de ítems de su dimensión, no fueron ubicados dentro de su grupo conceptual esperado. Estos ítems formaron un cluster independiente, y que requiere ser separado del contenido de su dimensión. Esto ocurrió en ambas áreas, factores de riesgo psicosocial (ítem 22 de Retroalimentación negativa del supervisor) y de Recursos Favorables (ítem 26 de Posibilidades de desarrollo laboral/profesional, e ítem 24 de Seguridad en la conservación del empleo). Parece claro que estos ítems que requieren recodificación adicionan varianza de método, debido al fraseo opuesto a la definición del constructo. Los hallazgos al respecto indican que el fraseo con orientación opuesta a su constructo introduce variabilidad no asociada al constructo, y en investigaciones asociadas a los factores psicosociales del trabajo y sus efectos, los contenidos fraseados opuestamente al constructo pueden producir diferente distanciamiento de las opciones de respuesta (Masuda, Kitaoka, & Ogino, 2012), un problema que cuestiona la presunción de medición intervalar igual de los ítems. El descuido en la respuesta también ha sido una posible explicación de la varianza introducida por este tipo de ítems de codificación inversa (Harris & Bladen, 1994). La evaluación de la equivalencia de

medición también tiende a encontrar a este tipo de ítems como fuente de variabilidad en los parámetros cuando de comparan grupos (Harris & Bladen, 1994). Para evitar que el modelo final del PROPSIT en la muestra peruana pueda introducir varianza irrelevante al constructo, es probable que las decisiones sobre los ítems de codificación inversa señalen su remoción.

Evaluación paramétrica

El objetivo de esta parte del estudio fue completar la evaluación de la dimensionalidad de los ítems del PROPSIT, mediante la implementación de metodología paramétrica en una muestra peruana diferente. Esta muestra fueron docentes de educación básica regular, y que comparten aproximadamente similar experiencia de interacción con la estructura organizativa, con el tipo de tareas asignadas, y con las personas involucradas en estas organizaciones educativas. Mediante metodología semi-confirmatorio de análisis factorial, se evaluó primero la configuración esperada de los ítems en sus dimensiones, y los resultados se evaluaron en la métrica del coeficiente de congruencia, y de la observación de la complejidad factorial de los ítems.

Respecto al área de Factores de Riesgo Psicosocial (FRPS), se halló inconsistente congruencia en los resultados, indicando que los ítems no se ajustaron satisfactoriamente en sus dimensiones esperadas, y algunas de estas dimensiones contenían ítems no esperados teóricamente. En ambas muestras separadas, la falta de congruencia general fue notoria (en cada muestra, la congruencia total fue $< .80$); sin embargo, se detectaron algunos patrones de relaciones ítem-factor consistentes entre las muestras, y en la muestra total. Este patrón fue evidenciado por los coeficientes de congruencia en el nivel de los ítems, pero también sustantivamente, porque su contenido era teóricamente explicable. Como consecuencia de lo anterior, se modificó el modelo de medición de FRPS; aunque esta modificación no es teóricamente óptima, ésta es una decisión metodológica habitual en la evaluación de modelos de medición, y en sus etapas de construcción y adaptación. En la segunda iteración del análisis, para la definición de las dimensiones se reubicaron los ítems con tres criterios: la racionalidad

teórica de las dimensiones propuestas del PROPSIT, los resultados empíricos de Juárez-García y Flores- Jiménez (2020), y la optimización de los factores mediante la reducción de la baja validez de los ítems y de su complejidad factorial. Incluido en estas decisiones, también se debieron remover ítems inestables en su ubicación dimensional, una decisión respaldada por el estudio de Juárez-García y Flores- Jiménez (2020), donde parcialmente removieron los mismos los ítems de los presentes resultados.

Los factores que emergieron con alta congruencia factorial fueron *demandas psicológicas* (F1), *exigencias físicas* (F3), y *acoso psicológico* (F4); es decir, los factores F2, F4 y F5 en Juárez-García y Flores-Jiménez. El factor que se mantuvo independiente, y que agregó una dimensión adicional a este modelo de FRPS, fue el grupo de ítems que vinculan la interacción con las personas y las reacciones emocionales negativas, el mismo que parece configurar un contenido de *demandas emocionales* (F2). En Juárez-García y Flores-Jiménez únicamente el ítem 12 (emociones negativas de otras personas) fue incluido en demandas psicológicas, mientras el resto de los ítems fueron removidos. Sin embargo, este mismo ítem mostró la carga factorial baja y moderada complejidad en sus análisis exploratorios, lo que puede sugerir potencial modificación en nuevos estudios.

En algunos de los ítems, la complejidad factorial fue clara, expresada cargas factoriales moderadamente altas en más de un factor, pero esto no representa un problema en términos absolutos; en cambio, para tomar decisiones sobre este tipo de estructura, se requiere ubicarse en un ángulo relativo para hallar el sentido del mismo. Por ejemplo, desde un ángulo empírico, la complejidad factorial de algunos ítems plantea que estas relaciones entre estos ítems y otras dimensiones son reales, y a menos estos ítems representen variaciones aleatorias o sesgo sistemático (eg., respuestas con esfuerzo insuficiente), la interpretación de estas relaciones se orienta hacia una posible nueva manera de interpretar las dimensiones. Por ejemplo, respecto a Recursos Favorables, en la muestra total (Tabla 6, encabezado Muestra total), el ítem 28 (Trabajo agradable y gratificante) representaba teóricamente al

factor recompensas y recursos, pero empíricamente se encuentra asociado a los aspectos del trabajo que facilitan el control del trabajo sobre su desempeño y uso de habilidades (ie, factor F2). Este ítem también muestra una relación no trivial con el factor que describe el apoyo social (Tabla 6, factor F4), algo que es factorialmente más claro en la muestra de educadores. En conjunto, esto sugiere la covariabilidad de los efectos positivos del intercambio social que da soporte al trabajador durante su trabajo diario (Nahum-Shani, Bamberger, & Bacharach, 2011). Aparentemente, esta relación está condicionada por el tipo de puesto, porque es más claro en la muestra de educadores comparada con la muestra general.

La limitación principal de este estudio fue la heterogénea distribución de una de las muestras, respecto a las carreras y puestos laborales de una de las muestras, y en que las conclusiones sobre su validación externa están comprometidas. Sin embargo, el tamaño del efecto de los resultados finales (ie, cargas factoriales, estimaciones de confiabilidad, baja complejidad factorial, y altos coeficientes de congruencia) pueden justificar la potencial generalización, y es un criterio no trivial para esta conclusión (Coppock, 2018; Coppock, Leeper, & Mullinix, 2018; Mullinix, Leeper, Druckman, & Freese, 2015). Específicamente, con cargas factoriales $> .50$ consideradas como altas (Ximénez, 2006), la versión final del PROPSIT puede alcanzar la alta validez interna y la replicabilidad de cada factor.

Una vez definida la dimensionalidad del PROSIT en la muestra analizada, la investigación sobre las relaciones entre sus dimensiones puede proceder y extenderse con rigurosidad.

3.6. Objetivo 3b: Obtener evidencias de validez convergente y divergente interna del PROPSIT.

3.6.1. Introducción

En el análisis del objetivo correspondiente con la dimensionalidad del PROPSIT (Objetivo 3a: Identificar la estructura dimensional del PROPSIT), en algunos de los ítems la complejidad factorial fue claramente expresada con cargas factoriales moderadamente altas en más de un factor. Esto representa un problema cuando se espera que el instrumento represente constructos factorialmente simples (Brown, 2015); es decir, que los ítems tengan una fuerte carga factorial en una dimensión y cargas alrededor de cero en otros factores (Fleming & Merino-Soto, 2005).

En contraste con lo anterior, la diferenciación psicométrica entre constructos señala la importancia de la independencia conceptual de ellos, y esta es una propiedad de la estructura interna de una medida (Bagozzi & Phillips, 1991; AERA et al., 2014). Como características dentro del modelo de medición, la estructura interna puede ser evaluada mediante la validez convergente y divergente entre los factores del instrumento, comparando la fuerza o variabilidad interna de los factores frente a la fuerza o covariabilidad entre los factores (Bagozzi & Phillips, 1991; Brown, 2015). Esta fuerza interna habitualmente se extrae de las cargas factoriales luego de hacer el modelamiento con métodos exploratorios o confirmatorios (Brown, 2015). Las cargas factoriales grandes producen mayor varianza retenida dentro de su dimensión, y asegura una fuerte validez convergente del modelo de medición (Fornell & Larcker, 1981; Bagozzi & Phillips, 1991).

Dentro de este marco, la independencia entre los constructos se expresa en reducidos montos de varianza compartida ellos. De este modo, en los estudios de

construcción y adaptación de medidas, y particularmente en los factores psicosociales del trabajo, la diferenciación entre las dimensiones permite construir variables no redundantes de las condiciones del ambiente laboral. Como indicador de calidad psicométrica general de un instrumento, el reporte y el establecimiento del grado de distintividad de las dimensiones es importante y necesaria (Voorhees, Brady, Calantone, & Ramirez, 2016).

Por lo tanto, se requiere establecer la distintividad de las dimensiones con métodos formales, para afirmar que no son colineales, y que representan distintos constructos interpretables (Bagozzi & Phillips, 1991; Franke & Sarstedt, 2019). Aunque es difícil establecer unívocamente un límite desde el cual separar la redundancia y la independencia empírica entre los constructos (Martínez-García & Martínez-Caro, 2009), como heurística general se acepta que altas correlaciones lineales cercanas a 1.0 son evidencias de la alta dependencia de los constructos, y que implican redundancia estadística entre ellas. Por otro lado, las correlaciones bajas o cercanas a cero indican alta independencia, lo que estadísticamente se conoce como ortogonalidad (Brown, 2015; Martínez-García & Martínez-Caro, 2009). Hay variedad de puntos de corte para separar la distintividad de la redundancia entre las correlaciones de las dimensiones, pero aparentemente la literatura se centra en por lo menos 3 criterios (Hilkenmeier, Bohndick, Bohndick, & Hilkenmeier, 2020) para distinguir el tamaño de las correlaciones entre las dimensiones: menor a 1 (Anderson & Gerbing, 1988; Bagozzi & Phillips, 1991, menor a .85 (Brown, 2015; Kline, 2011; Voorhees et al., 2016), y menor a .80 (Brown, 2015),

El presente objetivo evaluó la distintividad de las dimensiones halladas en el modelo revisado de las áreas de factores psicosociales del trabajo del PROPSIT, mediante la implementación de dos criterios de validez discriminativa de las dimensiones. Este objetivo de parte de la evaluación de la estructura interna (AERA et al., 2014), y

proporciona evidencia sobre la diferenciación de los constructos dentro de las áreas de factores riesgo psicosocial y factores de recursos favorables.

3.6.2. Método

Participantes

Los participantes fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), luego de remover los participantes con potenciales respuestas con insuficiente esfuerzo. El tamaño de la muestra fue 305 trabajadores peruanos, cuya base de datos fue limpiada por potenciales sesgos de respuesta (Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST).

Instrumentos

Escala de Procesos Psicosociales del Trabajo (Juárez-García, & Flores-Jiménez, 2020). El instrumento utilizado para el presente objetivo fue el obtenido del análisis de la dimensionalidad, y resuelto en el objetivo Objetivo 3a (Identificar la estructura dimensional del PROPSIT), mediante los métodos paramétricos. En los factores de riesgo psicosocial (FRPS), los factores establecidos fueron 4: demandas psicológicas (5 ítems), exigencias físicas (3 ítems), acoso psicológico (4 ítems), y demandas emocionales (3 ítems); en los factores recursos favorables, los factores establecidos fueron 3: recompensas y recursos (4 ítems), clima de apoyo social (5 ítems), y control laboral (5 ítems).

Procedimiento

Recolección de datos. Los participantes fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), y por lo tanto se aplica aquí los

mismos procedimientos de recolección y conformidad con la conducta ética como fueron implementados.

Análisis. Se utilizaron los resultados del modelamiento semi-confirmatorio expuestos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), en la Tabla 30 para los ítems de factores de riesgo psicosocial, y la Tabla 32 para los ítems de factores de Recursos Favorables. Desde estos resultados, se evaluó la validez discriminativa de las dimensiones empíricas del PROPSIT.

Validez discriminativa. Entre los criterios para determinar la validez discriminativa entre constructos dentro de un modelo de medición, el *criterio Fornell - Larcker* (1981; criterio F-L) y la *ratio heteroatributo – monoatributo* (HTMT: *Heterotrait-Monotrait Ratio*; Henseler, Ringle, & Sarstedt, 2015) son medidas relativamente simples para identificar correctamente si dos medidas son suficientemente discriminativas (Voorhees et al. 2016). Sin embargo, el criterio HTMT ha ganado más aceptación debido a su mayor sensibilidad (Cheung & Wang, 2017; Voorhees et al. 2016)

El HTMT compara las correlaciones inter-ítem de diferentes constructos (correlaciones heteroatributo – heterométrodo) con las correlaciones inter-ítems del mismo constructo (correlaciones monoatributo - heterométrodo). Las comparaciones son pareadas, y se obtiene un coeficiente HTMT para cada par de factores comparados. Este coeficiente se expresa como una ratio entre la media geométrica de ambas correlaciones medias. Para concluir que los factores son suficientemente diferenciados, hay varios puntos de corte para HTMT, en que el valor observado debe ser menor que el valor del punto de corte. Por regla general, HTMT debe ser menor a 1.0 (Bagozzi & Phillips, 1991), pero de acuerdo a Voorhees et al. (2016), el punto de corte más eficaz es $r < .85$, un valor que coincide con

Kline (2011) respecto al techo de la correlación para diferenciar la falta de discriminación entre medidas.

Para propósitos de comparación, se aplicó también el criterio Fornell-Larcker (criterio F-L), que consiste en comparar la varianza promedio extraída en un factor (AVE: *average variance extracted*), con la correlación observada (varianza compartida) de este factor con otro factor; esta comparación es pareada entre cada par de factores. El criterio de discriminación entre los constructos es que AVE (específicamente su raíz cuadrada) debe ser mayor a la correlación observada entre las dimensiones. De acuerdo a una simulación (Cheung & Wang, 2017), para la validez convergente se sugiere $AVE > .50$, añadido a cargas factoriales mayores a .50 en cada dimensión; y para validez divergente, las correlaciones interfactoriales deberían ser $< .70$.

3.6.3. Resultados

Criterio Fornell-Larcker. En la Tabla 33, todos los valores AVE son mayores a .75 en cada dimensión de factores de riesgo psicosocial (FRPS), y mayores a .81 en las dimensiones de factores de recursos favorables (RF). Estos valores se encuentran muy por encima del criterio de referencia ($AVE > .50$) para asegurar la validez convergente de cada constructo individual. Respecto a las correlaciones interfactoriales, en FRPS todas fueron menores a .70 (entre .31 y .56); en RF, las correlaciones fueron más elevadas que con las halladas en FRPS (entre .67 y .75), y solo la correlación entre F3 (Clima de apoyo social) y F1 (recompensas y recursos) fue mayor a .70. El resto de las correlaciones en RF estuvieron cercanamente debajo del criterio de discriminabilidad ($r > .70$).

Criterio HTMT. Los valores debajo de la diagonal con 1.0s (Tabla 33, encabezado criterio HTMT) son los correspondientes con este criterio. Se halló que todos los valores

debajo de la diagonal fueron menores a los diferentes criterios HTMT reportados en la literatura (ie., 1.0, .90, y .85), en FRPS y RF, incluido en la relación entre F1 y F3 de RF. Esta relación F1-F3 fue detectada como ausencia de discriminación entre los constructos mediante el criterio Fornell-Larcker.

Tabla 33

Validez discriminativa de los factores del PROPSIT

	Factores de Riesgo Psicosocial (FRPS)				Factores de Recursos Favorables (RF)			
	F1	F2	F3	F4	F1	F2	F3	
Criterio Fornell-Larcker					Criterio Fornell-Larcker			
FRSP1	.77				F1	.86		
FRSP2	.41	.72			F2	.69	.87	
FRSP3	.31	.52	.77		F3	.75	.67	.82
FRSP4	.37	.54	.56	.77	-	-	-	-
Criterio HTMT					Criterio HTMT			
FRF1	1				F1	1		
FRF2	.34	1			F2	.68	1	
FRF3	.30	.47	1		F3	.77	.69	1
FRF4	.26	.46	.52		-	-	-	-

Nota. En negrita: AVE = raíz cuadrada de average variance extracted (varianza promedio extraída). HTMT: heterotrait - monotrait

Factores de riesgo psicosocial: FRSP1 = Demandas psicológicas. FRSP2 = demandas emocionales. FRSP3 = exigencias físicas. FRSP4 = acoso psicológico. Factores de Recursos Favorables: FRF1 = recompensas y recursos. FRF2 = control laboral. FRF3 = Clima de apoyo social.

3.6.4. Discusión

En el presente objetivo, se añadió evidencia al estudio de la estructura interna del PROPSIT, mediante la evaluación de la validez convergente y validez divergente de su estructura factorial. El estudio de la estructura interna no solo incluye a la identificación de

los factores y los ítems distribuidos en ellos (AERA et al., 2014), sino también a la distintividad de las dimensiones. Se aplicaron dos criterios, uno clásico (criterio Fornell-Larcker) y un criterio relativamente nuevo (criterio HTMT), y se halló que, predominantemente, ambos criterios fueron cumplimentados por las correlaciones interfactoriales del PROPSIT; es decir, la varianza retenida en cada factor y la distinción psicométrica entre las mismas. Respecto al primero, la varianza retenida por cada factor fue suficientemente alta para garantizar la representatividad psicométrica de los ítems con sus constructos y la posible replicabilidad de los resultados (Hancock & Mueller, 2000). La fuerza de la calidad de la solución factorial, debida a la fuerza de las cargas factoriales, facilitar identificar la claridad de la estructura factorial del PROPSIT y la precisión de los puntajes para evaluaciones individuales (Hancock & Mueller, 2000). Sin embargo, esta afirmación sobre la precisión de los puntajes requiere estimar parámetros adicionales para estar seguros de esta conclusión, orientados hacia la estimación de la confiabilidad.

Segundo, respecto a la diferenciación de constructos, excepto la relación entre el factor F1 (Recompensas y recursos) y F3 (Clima de apoyo social), el resto cumplió satisfactoriamente ambos criterios utilizados. La relación F3-F1 únicamente no cumplió el criterio Fornell-Larcker, y debido que el criterio HTMT se elaboró como una propuesta más precisa y recomendada para tomar decisiones finales sobre la distintividad entre los constructos (Cheung & Wang, 2017; Franke & Sarstedt, 2019; Henseler et al., 2015), la relación entre F3 y F1 puede establecerse en discriminativa. Sin embargo, esta discrepancia sugiere adicional exploración con métodos adicionales que balanceen las desventajas y ventajas de cada uno (Rönkkö & Cho, 2020). Finalmente, se pueden requerir explorar la capacidad de cada dimensión para estar asociado con diferencialmente con criterios

externos, y conocer la relevancia de cada dimensión dentro de un marco de red nomológica con otros constructos.

Se concluye que la adaptación del PROPSIT en la muestra peruana cuenta con distintividad de los constructos con base teórica y empírica, y que ésta última fortalece la diferenciación conceptual para describir los factores psicosociales del trabajo, conocidos como de riesgo o como recursos favorables dentro del marco teórico del PROPSIT.

3.7. Objetivo 3c: Examinar la equivalencia de medición del PROPSIT

3.7.1. Introducción

La equivalencia de medición es una propiedad de las medidas que indica que, la estructura interna de un instrumento en un grupo, es similar en otro grupo, y éste es un componente de la estructura de una medida (AERA et al., 2014). Esta propiedad es importante porque la comparación entre grupos en el constructo medido está condicionada por el grado en que está igualmente calibrado para todos los grupos comparados (Byrne, 2016; Meredith, 1993). La igualdad absoluta de los aspectos calibrados es seguramente improbable, y por lo tanto en el contexto práctico es poco realista esperar o hallar resultados en un nivel absoluto de igualdad paramétrica entre grupos. Contrario a esto, es más razonable esperar una cercana equivalencia entre la estructura interna de los constructos. La propiedad de equivalencia es conocida como *invarianza de medición* (Meredith, 1993).

La nominación de “invarianza de medición” generalmente ocurre en el contexto del análisis del modelamiento de ecuaciones estructurales (Bentler & Dugeon, 1996; Meredith, 1993), o más exactamente *análisis factorial confirmatorio multigrupo* (*multigroup confirmatory factor analysis*, MGCFA). El MGCFA está asociado a la detección del funcionamiento diferencial de ítems (*differential item functioning*, DIF; Raju, Laffitte & Byrne, 2002; Teresi, Ramirez, Lai & Silver, 2008; Whitmore & Schumacker, 1999). DIF es un término asociado a la teoría de respuesta al ítem (*ítem response theory*, IRT; Teresi et al., 2008) para la investigación del sesgo de respuesta a los ítems. Un instrumento compuesto por ítems invariantes o sin DIF, significa que las diferencias en los puntajes

pueden ser interpretados como asociadas al constructo, y en menor medida a diferencias en las propiedades psicométricas (Teresi et al., 2008; te Nijenhuis & van der Flier, 2009).

Los estudios sobre factores psicosociales del trabajo requieren corroborar que las propiedades se mantienen similarmente calibradas entre los grupos de trabajadores, debido a motivos metodológicos y teóricos. Por motivos metodológicos, los instrumentos con propiedades no-invariantes muestran desigualdades en los parámetros básicos de respuesta a un ítem, como los interceptos y los coeficientes beta (ie, coeficientes de discriminación o validez de los ítems). Debido las diferencias en estos parámetros, las diferencias entre grupos pueden ser confundidas como diferencias en el constructo, es decir, como diferencias sustantivas que conducen a interpretaciones conceptuales (Meredith, 1993). Estas diferencias son mejor descritas como funcionamiento el *funcionamiento diferencial de ítems* entre los grupos comparados. Por otro lado, por motivos teóricos la invarianza es importante porque la definición de constructo medido se apoya en el contenido elegido en cada subescala y en sus parámetros estructurales, y por lo tanto la falta de equivalencia supone también la falta de consistencia en la definición del constructo. Esta inconsistencia teórica significa que los grupos comparados están interpretando distintamente los ítems, y en conjunto, la variable latente que produce estas respuestas a los ítems.

Con algunas excepciones (eg., Abdin et al., 2019; Chungkham, Ingre, Karasek, Westerlund, & Theorell, 2013), en recientes validaciones de medidas de factores psicosociales en diferentes países e idiomas (eg., Ferrario, Fornari, Borchini, Merluzzi, & Cesana, 2005; Jacobo, & Máynez, 2020; Hökerberg et al., 2010; Ren, Li, Yao, Pi, & Qi, 2019; Houdmont, Randall, Kerr, & Addley, 2013; Toderi, & Sarchielli, 2016; Uribe, Gutiérrez, & Amézquita, 2020; Wood et al., 2019), la invarianza de medición fue

parcialmente evaluada con coeficientes de congruencia, o no fue evaluada completamente; y en general, apenas es discutido en las diferentes revisiones sobre instrumentos de factores psicosociales (Magnavita, 2008; Tabanelli et al., 2008). Más aún, en estudios que incluyen dos grupos culturalmente distintos (Hahr et al., 2014), o altamente heterogéneos (Berthelsen, Westerlund, Bergström, & Burr, 2020) la invarianza no es evaluada, y pueden suponerse dudas sobre la validez de sus comparaciones entre-grupos.

La necesidad de verificar la invariabilidad de las propiedades psicométricas como fuente de posible sesgo en las diferencias constructo en diferentes grupos, ha sido resaltada como una potencial limitación que debe requerir más atención en la medición de los factores psicosociales del trabajo (Orhede & Kreiner, 2000). Esta ausencia posiblemente ocurre porque la definición de validez de constructo implementada en estos estudios no está actualizada hacia visión integrativa de las diferentes facetas de la validez, donde la equivalencia de medición es un componente definicional de la estructura interna de una medida (AERA et al., 2014), y se recurre a inducir la validez de los instrumentos desde otros estudios, un problema que ha sido resaltado en otras medidas de diferentes constructos (Merino-Soto & Angulo-Ramos, 2020, 2021; Merino-Soto, & Calderón-De la Cruz, 2018).

Entre las decisiones que deben ser tomadas en un análisis de invarianza de medición, aquí resaltaremos dos. Una de las primeras decisiones en aplicar el modelamiento SEM para validación de instrumentos es el estimador estadístico que se usará para evaluar el ajuste del modelo de medición propuesto (Byrne, 2016; Sellbom, & Tellegen, 2019). Este estimador sirve para producir un estadístico de tipo χ^2 , que proviene de la comparación entre el modelo propuesto, y el modelo estimado estadísticamente.

Algún estimador requiere el cumplimiento de algún grado de presunciones respecto a la normalidad multivariada de las variables observables, así como la forma de registro de la variable observada o escalamiento, esto es, si es continua o categórica (Muthén, du Toit, & Spisic, 1997; Savalei, 2014; Sellbom, & Tellegen, 2019). El estimador apropiado cuando los datos cumplen la normalidad multivariada es de *máxima verosimilitud* (*maximum likelihood*; Savalei, 2014), pero en caso contrario (como es habitual en la investigación aplicada), existen otros estimadores más precisos y aplicables al análisis de variables categóricas, como ocurre en los ítems ordinales. Estos métodos hacen correcciones a la prueba estadística χ^2 , y al error estándar de los parámetros calculados (eg., cargas factoriales), para obtener resultados robustos (Muthén, 1993; Satorra, & Bentler, 1994). Uno de los estimadores recomendados para este el modelamiento de variables categóricas es el *weighted least square mean and variance adjusted* (WLSMV, Muthén et al., 1997), que usa el vínculo probit para definir la relación funcional entre los ítems y el constructo, así como correlaciones policóricas entre los ítems para estimar el grado de asociación lineal entre los ítems y para obtener umbrales (*thresholds*) entre categorías de respuestas.

Otra de las decisiones para implementar un análisis de comparación entre-grupos es el balance de las muestras comparadas, es decir, que las características basales de los grupos sean iguales o muy similares. Esto puede ser particularmente difícil de alcanzar en los estudios no experimentales y transversales, en que la potencia estadística se maximiza con la obtención del mayor número posible de participantes, y la representatividad poblacional no está garantizada. Como tendencia general, el muestreo por conveniencia tiene una alta prevalencia (Winton, & Sabol, 2021; Sarstedt, Bengart, Shaltoni, & Lehmann, 2018; Scholtz, de Klerk, & de Beer, 2020), y es un método cuya elección influye

su rapidez, costos, proximidad y accesibilidad (Gravetter, & Forzano, 2012, pp. 131-152). Es razonable pensar en la limitada propiedad del muestreo de conveniencia para la generalización de resultados; sin embargo, en los parámetros de medición, hay una consistencia de resultados aún entre diferentes muestras de conveniencia (Winton, & Sabol, 2021), compuestas de variadas características. Sin embargo, esto puede depender del contexto de investigación, y hay una necesidad de maximizar el grado de igualdad entre las muestras (Ho, Imai, King, & Stuart, 2007, 2011; Olmos, & Govindasamy, 2015).

El objetivo del presente trabajo fue evaluar la invarianza de la estructura interna del PROPSIT, entre trabajadores peruanos y mexicanos. La comparabilidad de las muestras de ambos grupos nacionales fue maximizada con la implementación de puntajes de propensividad, un método para reducir la influencia de potenciales covariables en la estimación de las diferencias entre grupos, que en este estudio fue la evaluación de la invarianza de medición.

3.7.2. Método

Participantes

Muestra peruana. Los participantes fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), luego de remover los participantes con potenciales respuestas con insuficiente esfuerzo. El tamaño de la muestra fue 305, luego que fue limpiada de casos por potenciales respuestas sesgadas (ie., respuestas descuidadas o con insuficiente esfuerzo; ver Objetivo 2: Identificar sesgos en las respuestas al PROPSIT – FPST. La distribución de las principales características sociodemográficas, antes del procedimiento de emparejamiento, aparece en la Tabla 34, bajo el encabezado Muestras no emparejadas.

Muestra mexicana. Fueron 409 trabajadores operativos de una empresa mexicana en Cuernavaca, Morelos, para la elaboración de productos utilizados en las prestaciones de salud. La empresa fue elegida por accesibilidad y proximidad al Observatorio de Factores Psicosociales y Bienestar Organizacional, dependencia del Centro de Investigación Transdisciplinar en Psicología (CITPSI, Universidad Autónoma del Estado de Morelos). Una parte de esta muestra, la correspondiente con una la empresa elegida, fue analizada en el estudio de validación del PROPSIT en trabajadores mexicanos de manufactura (Juárez-García, & Flores-Jiménez, 2020). La distribución de las principales características sociodemográficas, antes del procedimiento de emparejamiento, aparece en la Tabla 34, bajo el encabezado Muestras no emparejadas.

Instrumentos

Escala de Procesos Psicosociales del Trabajo (Juárez-García, & Flores-Jiménez, 2020). El instrumento utilizado para el presente objetivo fue el obtenido del análisis de la dimensionalidad, y resuelto en el Objetivo 3a: Identificar la estructura dimensional del PROPSIT, mediante los métodos paramétricos. En los factores de riesgo psicosocial (FRPS), los factores establecidos fueron 4: demandas psicológicas (5 ítems), exigencias físicas (3 ítems), acoso psicológico (4 ítems), y demandas emocionales (3 ítems); en los factores de recursos favorables (FRF), los factores establecidos fueron 3: recompensas y recursos (4 ítems), clima de apoyo social (5 ítems), y control laboral (5 ítems).

Procedimiento

Recolección de datos. Los participantes peruanos fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), y por lo tanto se aplica aquí los

mismos procedimientos de recolección y conformidad con la conducta ética como fueron implementados. En la muestra de trabajadores mexicanos, el procedimiento fue descrito en el estudio de validación de la estructura interna, de Juárez-García y Flores-Jiménez (2020)

Emparejamiento de las muestras. Debido a la posible presencia de sesgo de selección en las muestras no probabilísticas en la investigación de diseño transversal (Winton, & Sabol, 2021; Sellbom, & Tellegen, 2019), se implementó una estrategia de igualación de muestras en variables consideradas clave (o covariables) y consistentes. Para lograr la comparabilidad de las muestras, se obtuvieron los puntajes de propensividad (puntajes PS), calculados mediante regresión logística con función de vínculo logit en que el criterio predicho es el estatus de pertenencia del participante a unos de los grupos comparados (esto es, trabajadores peruanos o trabajadores mexicanos), y los predictores (o covariables) fueron los seleccionados por su consistencia (ver más adelante). El método de emparejamiento fue el *vecino más cercano* (Nearest Neighbor), en que el tamaño muestral de referencia o target fue la muestra peruana. Una vez obtenidos los puntajes PS, cada participante de cada grupo fue comparado en base a este puntaje, y se identificó los pares de PS que son similares. Por lo tanto, los participantes emparejados, es decir, que hacen “match”, son aquellos en que la probabilidad del PS es igual o muy similar, mientras que los PS más distantes fueron considerados no emparejados. La comparación fue *1:1*, es decir, entre un participante con otro participante de igual o muy cercano PS. El emparejamiento terminó cuando el tamaño de la muestra de comparación fue igual a la muestra target (muestra peruana, $n = 305$). Se usaron los programas R *MachIt* (Ho, Imai, King, & Stuart, 2011), y *tableone* (Yoshida, K., & Bartel, 2021).

Existieron covariables comunes y no comunes en ambas bases de datos, pero algunas de estas variables tuvieron codificación diferente respecto al puesto laboral, el área

de trabajo, o la carrera de estudio. Las covariables elegidas para generar los PS fueron las que su registro fue unívoco, y sin alguna incertidumbre en su interpretación. De este modo, las covariables elegidas fueron la edad (registrada en años), el estado civil y el grado de instrucción. Estas covariables fueron también reportadas en el estudio de validación mexicana del PROPSIT (Juárez-García, & Flores-Jiménez, 2020), y también están identificados como características sociodemográficas clave en otros estudios de validación intercultural (eg., Hahr et al., 2014). El sexo no fue considerado debido al extremo desbalance en la muestra mexicana (hombres, $n = 350$, 85.6%; mujeres, $n = 59$, 14.4%).

Análisis estructural. Como fuente principal de variabilidad en las propiedades psicométricas, se eligió la agrupación de los trabajadores en mexicanos y peruanos. Esta agrupación tuvo como marco las diferencias interculturales entre los constructos (Meredith, 1993), con el presupuesto de las posibles diferencias en la estructura interna del PROPSIT vinculadas con la nacionalidad de los trabajadores.

Se aplicó un análisis factorial confirmatorio (Jöreskog, 1969) desde el modelamiento de ecuaciones estructurales (*Structural Equation Modeling, SEM*; Bentler & Dugeon, 1996), para a) verificar la dimensionalidad de los ítems en la muestra mexicana y peruana, y b) para posteriormente la invarianza de medición. La función estadística elegida estuvo basada en la metodología para variables categóricas, específicamente *weighted least square mean and variance adjusted* (WLSMV; Muthén, du Toit, & Spisic, 1997), un estimador que tiende a ser efectivo para obtener parámetros más exactos cuando las variables son categóricas ordinales y no-normales (Finney & DiStefano, 2013; Li, 2016), particularmente con fuerte asimetría (Sass, Schmitt & Marsh, 2014). También, tiende a mostrar consistencia en sus estimaciones en una variedad de situaciones (Yang-Wallentin, Jöreskog, & Luo, 2010).

El ajuste del modelo de medición del PROPSIT fue evaluado mediante indicadores prácticos recomendados por la literatura metodológica (Jackson, Gillaspay & Purc-Stephenson, 2009), en dos niveles de ajuste (Maydeu-Olivares 2017): cercano ajuste ($RMSEA \leq .05$, $SRMR \leq .05$, and $CFI \geq .95$; Maydeu-Olivares 2017), y razonable ajuste ($RMSEA \leq .08$, $SRMR \leq .08$, and $CFI \geq .90$ (Hu & Bentler 1999). Alguna modificación post hoc será evaluada examinando los *índices de modificación* (IM; Sorbom, 1989), y la intensidad de su efecto sobre el cambio del modelo de medición. Para esto último, se aplicó el método de Saris, Satorra y van der Veld (2009), en que se especifica a priori un valor mínimo (no trivial) para el parámetro de interés (e.g., carga factorial, error correlacionado, etc.) y una potencia estadística. Para el presente estudio, el nivel mínimo para los errores correlacionados fue .20, con una potencia estadística de .80. Estos posibles ajustes al modelo también evaluados su relevancia conceptual, para reducir la dependencia de estas decisiones a los resultados estadísticos que también pueden estar influenciados por variaciones idiosincrásicas de las muestras de participantes recolectadas (Byrne, 2008).

La invarianza de medición se hizo con un enfoque bottom-up, desde un modelo sin restricciones hasta un modelo con fuertes restricciones (Meredith, 1993; Meredith & Teresi, 2006; Stark, Chernyshenko, & Drasgow, 2006). De este modo, fueron probados: un modelo sin restricciones de igualdad (invarianza configuracional) y continuamos con restricciones sucesivas aplicadas a las cargas factoriales y umbrales (invarianza métrica), e interceptos (invarianza escalar). Tomando en cuenta el tamaño muestral (> 300 ; Chen, 2007), el criterio de invarianza fue: $\Delta_{CFI} < .010$, $\Delta_{SRMR} < .030$, y $\Delta_{RMSEA} < .015$ (Chen, 2007). Los programas R utilizados fueron *lavaan* (Rosseel, 2012) y *semTools* (Jorgensen, Pornprasertmanit, Schoemann, & Rosseel, 2018).

3.7.3. Resultados

Emparejamiento entre las muestras

En la Tabla 34, debajo del encabezado Muestras emparejadas, aparece la distribución de las covariables elegidas luego de la selección por puntajes PS. En general, las diferencias entre ambas muestras emparejadas fueron pequeñas, como se ve en los coeficientes V-Cramer (diferencias en el estado civil), d-Cohen (diferencias en edad) y γ (en el grado de instrucción). En cambio, en las muestras no emparejadas, las diferencias fueron mayores. En un análisis más detallado, las diferencias entre las proporciones de cada categoría fueron evaluadas en la métrica de diferencias arco seno (coeficiente h ; Cohen, 2008). Estas fueron predominantemente pequeñas ($< .50$) o triviales ($< .20$), y todas disminuyeron comparadas con las diferencias en las muestras antes del emparejamiento. En la Figura 16 se muestra la distribución de los puntajes de propensividad en la antes y después de aplicar el procedimiento de emparejamiento, para las muestras comparadas.

Tabla 34

Balance entre las muestras no emparejadas y emparejadas

	Muestras no emparejadas (n total = 714)						Muestras emparejadas (n total = 610)					
	Muestra mexicana N = 409		Muestra peruana N = 305		Evaluación del balance		Muestra mexicana (n = 305)		Muestra peruana (N = 305)		Evaluación del balance	
	M	DE	M	DE	Prueba estadística	Tamaño del efecto	M	DE	M	DE	Prueba estadística	Tamaño del efecto
Edad	34.6	9.4	35.1	11.2			34.1	9.2	35.1	11.2		
Estado civil	N	%	N	%	$\chi^2 = 35.1$	$V = .22$ h	N	%	N	%	$\chi^2 = 6.54$	$V = .104$ h
Soltero(a)	138	33.7	168	55.1	-	-0.43	138	45.2	168	55.1	-	-.19
Casado(a)	201	49.1	98	32.1	-	0.34	124	40.7	98	32.1	-	.17
Unión libre	57	13.9	36	11.8	-	0.06	38	12.5	36	11.8	-	.02
Divorciado(a)	12	2.9	3	1.0	-	0.14	5	01.6	3	01.0	-	.05
Viudo(a)	1	0.2	0	0.0	-	0.09	-	-	-	-	-	-
Grado de instrucción	N	%	N	%	$MH-\chi^2 = 9.06$	$\gamma = .112$ h	N	%	N	%	$MH-\chi^2 = 6.35$	$\gamma = .00$ h
Primaria terminada	6	1.5	0	0.0	-	0.24	-	-	-	-	-	-
Secundaria terminada	82	20.0	35	11.5	-	0.23	41	13.4	35	11.5	-	.06
Bachillerato terminado/ carrera técnica	122	29.8	138	45.2	-	-0.32	121	39.7	138	45.2	-	-.11
Licenciatura/ Universidad terminada	189	46.2	82	26.9	-	0.40	133	43.6	82	26.9	-	.35
Posgrado terminado	10	2.4	50	16.4	-	-0.52	10	3.3	50	16.4	-	-.52

Nota. V = coeficiente V-Cramer. γ : coeficiente gamma. χ^2 : prueba de independencia entre nominales. $MH-\chi^2$: prueba Mantel-Haenszel de independencia entre variables ordinales. h: diferencia entre transformaciones arcoseno de proporciones.

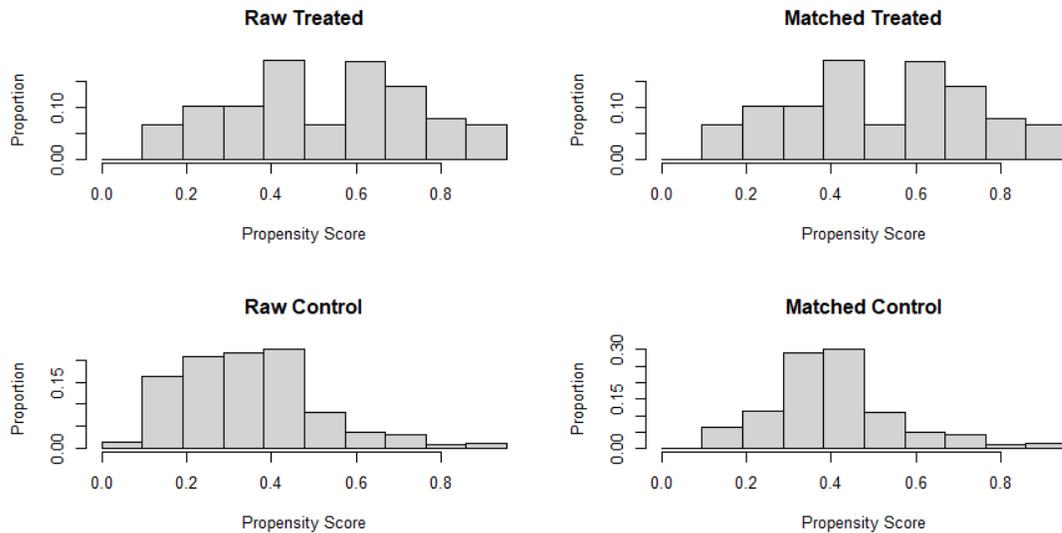


Figura 16
Puntajes de propensividad antes y después del emparejamiento

Invarianza de medición

Factores de Riesgo Psicosocial. Primero, fue evaluada la configuración (eg, número de dimensiones) en ambas muestras, cuyos resultados aparecen debajo del encabezado “En cada grupo” de la Tabla 35. En la muestra mexicana, el ajuste de las dimensiones fue satisfactorio, donde todos los coeficientes de ajuste fueron aproximadamente satisfactorios. Pero en la muestra peruana, el ajuste fue parcialmente satisfactorio, debido que los índices RMSEA y SRMR no cumplieron con el criterio de ajuste.

Luego de esta evaluación, los resultados de la invarianza se obtuvieron respecto a la invarianza configuracional, métrica y escalar (Tabla 35). Los resultados indican que el

cambio en el ajuste entre los modelos comparados fue pequeño, y consistentemente menores comparados con el criterio elegido de Chen (2007). Considerando la muestra total, el ajuste fue aproximadamente satisfactorio, WLSMV - $\chi^2 = 322.90$ (gl = 84), CFI = .991, RMSEA = .068 (IC 90% = .061, .076), SRMR = .060. Los parámetros (cargas factoriales y correlaciones interfactoriales en la muestra total aparecen en la Tabla 36.

Factores de Recursos Favorables. En cada una de las muestras comparadas (Tabla 35, encabezado “En cada grupo”), el ajuste puede considerarse aceptable, aunque existió discrepancia con el índice RSMEA. En la evaluación del cambio de ajuste desde la invarianza configuracional a la invarianza escalar, el cambio fue pequeño, excepto el cambio del RMSEA. Considerando la muestra total, el ajuste fue moderadamente satisfactorio, WLSMV - $\chi^2 = 392.51$ (gl = 74), CFI = .990, RMSEA = .084 (IC 90% = .076, .092), SRMR = .052; los parámetros (cargas factoriales y correlaciones interfactoriales) en la muestra total aparecen en la Tabla 36.

3.7.4. Discusión

El objetivo del presente estudio fue analizar la invarianza de medición del PROPSIT entre trabajadores una muestra peruana y otra mexicana. Debido al desbalance o heterogeneidad de los grupos comparados en sus ocupaciones laborales y áreas de carrera, ambas muestras fueron emparejadas hacia la máxima similaridad con una metodología aceptada para homogeneizarlas, o hacerlas similares: el puntaje de propensividad. Este procedimiento ayudó a controlar la posible variabilidad derivada de fuentes diferentes al estatus de origen de los grupos (Ho et al., 2007, 2011; Olmos & Govindasamy, 2020): en este estudio, el grado de instrucción, la edad y el estado civil.

Tabla 35

Ajuste del modelo de medición e invarianza de medición

	Ajuste de los modelos				Cambio en el ajuste		
	WLSMV χ^2 (gl)	CFI	RMSEA	SRMR	Δ_{CFI} ($< .010$)	Δ_{RMSEA} ($< .015$)	Δ_{SRMR} ($< .030$)
F. Riesgo Psicosocial							
En cada grupo							
Perú	339.91 (84)	.980	.100 (.08, .11)	.102	-	-	-
México	105.79 (84)	.994	.053 (.039, .066)	.056	-	-	-
Invarianza							
Configuración	494.79 (168)	.987	.080 (.072, .088)	.079	-	-	-
Métrica	789.64 (234)	.977	.088 (.082, .095)	.086	.003	.002	.007
Interceptos	883.70 (245)	.974	.093 (.086, .099)	.083	.003	.011	.003
F. Recursos Favorables							
En cada grupo							
Perú	258.21 (74)	.997	.090 (.079, .103)	.049	-	-	-
México	244.18 (74)	.979	.087 (.075, .099)	.069	-	-	-
Invarianza							
Configuración	504.71 (148)	.995	.089 (.081, .098)	.059	-	-	-
Métrica	1151.926 (214)	.987	.120 (.113, .127)	.080	.008	.031	.021
Interceptos	929.83 (225)	.991	.102 (.095, .108)	.062	.004	.018	.018

Tabla 36

Análisis factorial confirmatorio: muestras nacionales y muestra total

	Muestra peruana (n = 305)				Muestra mexicana (n = 305)				Muestra total (n = 710)			
	FRPS1	FRPS2	FRPS3	FRPS4	FRPS1	FRPS2	FRPS3	FRPS4	FRPS1	FRPS2	FRPS3	FRPS4
F. Riesgo Psicosocial												
PROP1	.869				.752				.836			
PROP2	.904				.771				.858			
PROP6	.688				.645				.726			
PROP10	.826				.739				.791			
PROP11	.593								.664			
PROP12		.698				.833				.805		
PROP13		.901				.917				.924		
PROP14		.718				.823				.759		
PROP15			.800				.736				.794	
PROP16			.986				.939				.934	
PROP17			.568				.567				.657	
PROP18				.954				.859				.891
PROP19				.877				.897				.890
PROP20				.844				.829				.853
PROP21				.854				.648				.733
Correlaciones												
FRPS1	1				1				1			
FRPS2	.411	1			.721	1			.623	1		
FRPS3	.313	.523	1		.471	.488	1		.485	.556	1	
FRPS4	.378	.544	.567	1	.592	.744	.561	1	.549	.707	.615	1
F. Recursos Favorables												
	FRF1	FRF2	FRF3		FRF1	FRF2	FRF3		FRF1	FRF2	FRF3	
PROP24	.692			-	.490			-	.587			-
PROP25	.845			-	.585			-	.694			-
PROP35	.874			-	.613			-	.737			-
PROP36	.837			-	.635			-	.733			-
PROP30		.815		-		.571		-		.692		-
PROP31		.970		-		.824		-		.907		-
PROP32		.975		-		.910		-		.934		-
PROP33		.879		-		.242		-		.556		-
PROP34		.920		-		.550		-		.717		-
PROP37			.803	-			.671	-			.720	-
PROP38			.836	-			.756	-			.781	-
PROP39			.857	-			.785	-			.821	-
PROP40			.949	-			.811	-			.864	-
PROP41			.875	-			.769	-			.805	-
Correlaciones												
FRF1	1			-	1			-				-
FRF2	.696	1		-	.657	1		-				-
FRF3	.750	.679	1	-	.739	.697	1	-				-

Nota. Factores de riesgo psicosocial: FRSP1 = Demandas psicológicas. FRSP2 = demandas emocionales. FRSP3 = exigencias físicas. FRSP4 = acoso psicológico. Factores de Recursos Favorables: FRF1 = recompensas y recursos. FRF2 = control laboral. FRF3 = Clima de apoyo social.

Discusión

El balance conseguido no fue perfectamente exacto, pero el nivel de similaridad alcanzado debe evaluarse en el contexto del poder estadístico y la estabilidad de los resultados obtenidos. Es decir, que un emparejamiento más exacto puede obtenerse con el costo de un menor tamaño muestral, en que los casos emparejados mostrarían muy cercanos puntajes de propensividad, y se desecharían los casos con puntajes de propensividad con moderadas, pero aún pequeñas diferencias. Esto conduce a reducir el tamaño muestral, la potencia estadística, y la estabilidad de los parámetros psicométricos de interés (i.e., número de dimensiones, cargas factoriales e interceptos). Por lo tanto, el emparejamiento de las muestras fue aceptable para el presente objetivo. La aplicación de esta metodología de puntajes de propensividad es una fuerza del presente estudio, porque las inferencias sobre la equivalencia del PROPSIT pueden dirigirse con mayor validez en la variabilidad producida por el grupo de interés (i.e., nacionalidad de los trabajadores: peruanos vs mexicanos), mientras que se reducen los posibles efectos de las diferencias de los examinados en otras variables (en este estudio, edad, grado de instrucción y estado civil). Junto con el sexo de los participantes, estas dos variables fueron importantes marcadores descriptivos en el estudio de validación psicométrica del PROPSIT en trabajadores mexicanos (Juárez-García & Flores-Jiménez, 2020).

Respecto a la configuración, o número de dimensiones, en los factores de riesgo psicosocial fue aceptable en la muestra mexicana, y en la muestra peruana cercanamente aceptable. Esto sugiere que el número de dimensiones en la muestra peruana puede no ser exacta como es representada en el PROPSIT, y contener más o menos dimensiones. Sin

embargo, dado que en el Objetivo 3a (Identificar la estructura dimensional del PROPSIT) del presente proyecto se exploró a fondo esto y se concluyó óptimamente sobre el número de dimensiones del PROPSIT, entonces el bajo ajuste comparativo de la configuración factorial (e.g., número de dimensiones) del PROPSIT en la muestra peruana puede ser consecuencia de otros aspectos. Uno de estos puede ser el enfoque factorial elegido para evaluar el ajuste, es decir, análisis factorial confirmatorio (CFA; Jöreskog, 1969). Este enfoque, aunque es muy popular entre los estudios de validez de la estructura interna de instrumentos de factores psicosociales, puede no representar con exactitud las relaciones entre los ítems y sus factores. El CFA estima las cargas factoriales de los ítems en sus factores hipotetizados, pero considera que las cargas factoriales en los otros factores es cero. Si esto no es así, entonces hay un decremento en el grado de ajuste (Asparouhov, & Muthén, 2009).

Sobre la invarianza de las cargas factoriales, una implicación importante es la definición del constructo son similares entre las muestras comparadas. Las cargas factoriales ayudan a delimitar empíricamente la definición de constructo, porque ellas representan la fuerza de esta relación (Meredith & Teresi, 2006). Cuando las muestras comparadas contienen similares cargas factoriales, las diferencias pueden considerarse dentro del error de muestreo, conceptualmente indica también la similaridad con que el constructo es establecido en cada muestra. Una implicación es que, aún con las diferencias naturales en la aplicación de los instrumentos, y las diferencias fenotípicas y psicosociales de ambas muestras, la conceptualización de las dimensiones puede ser similar.

La invarianza escalar mostró ser satisfactoria en las dimensiones del área Factores de Riesgo Psicosocial, pero fue sospechosa en el los Factores de Recursos Favorables. Este

aparente cuestionamiento del ajuste provino del índice RMSEA, pero no de CFI y SRMR. Esto sugiere los interceptos pueden ser diferentes, y no sería viable la comparación entre-grupos respecto a los puntajes promedios de las dimensiones. Sin embargo, esta discrepancia puede ser de magnitud pequeña, y estar incrementadas por el tipo de modelamiento utilizado, es decir, CFA. Debido que solo el índice RSMEA es discrepante, provisionalmente se puede aceptar que la invarianza escalar está cumplida.

Sobre las limitaciones del estudio: una de las posibles limitaciones fue la necesidad de reducir las opciones de respuesta, de 7 a 5 categorías, en 5 ítems del área de factores de riesgo psicosocial (ítems de Acoso psicológico: 16, 18, 19, 20, 21) en la muestra peruana, y de un ítem en la dimensión Clima de Apoyo Social (ítem 40) en la muestra mexicana. Estas categorías de respuesta combinadas fueron las ultimas opciones de respuesta, en que se describe la alta intensidad del acoso recibido. El efecto de esta reducción en una parte de los ítems puede tener una influencia en el ajuste, produciendo resultados óptimos pero que son artefactos metodológicos (Rutkowski, Svetina, & Liaw, 2019). Sin embargo, el número de ítems modificados fue pequeño comparado con el número total de ítems no modificados, y probablemente su influencia sobre el ajuste puede considerarse trivial. Otra limitación fue que el modelamiento de la estructura del PROPSIT puede requerir la estimación directa de las cargas factoriales cruzadas, como ocurre en el análisis factorial exploratorio o en el más reciente exploratory structural equation modeling (ESEM; Asparouhov, & Muthén, 2009).

En el modelamiento ESEM, se estiman las cargas factoriales de los ítems en su dimensión hipotetizada y en las dimensiones no hipotetizadas. Esto crea una importante diferencia con el análisis factorial confirmatorio (CFA; Jöreskog, 1969), en que únicamente las cargas factoriales de los ítems en su dimensión son estimadas. Esta diferencia se ve en

el grado de ajuste, porque los modelos ESEM siempre se ajustan mejor que los modelos CFA. Manteniendo el resto de cosas iguales, este superior ajuste proviene porque las cargas cruzadas realísticamente muestran montos de varianza que requieren ser modeladas (Asparouhov, & Muthén, 2009). En el presente estudio, el ajuste de la configuración (o número de factores) del PROPSIT fue bueno en la muestra mexicana, y moderadamente aceptable en la muestra peruana, y estos niveles de ajuste pueden estar subestimados porque el CFA no estimó las cargas cruzadas. Del mismo modo, el grado de covariación entre las dimensiones pueden también estar infladas por el mismo motivo (Asparouhov, & Muthén, 2009). En el futuro, este modelamiento ESEM requiere ser implementado para dar más exactitud a la estructura interna del PROPSIT, pero en una muestra mayor. En relación a lo anterior, una tercera limitación fue el tamaño de las muestras comparadas, que fue relativamente pequeña (alrededor de 300) y no permitió examinar la invarianza dentro de cada una, por ejemplo, de acuerdo al grado de instrucción. Por lo tanto, un tamaño muestral mayor permitirá dividir su tamaño en los grupos comparados y examinar la invarianza intra-muestra dentro de cada grupo nacional (por ejemplo, ver Juárez García, Merino Soto, Fernández, Flores Jiménez, Caraballo, & Camacho Cristiá, 2020), como un pre-procesamiento del análisis de invarianza entre los grupos nacionales. Finalmente, la comparación entre diferentes grupos de trabajadores puede ser una limitación para la generalizar los resultados a otros grupos laborales. Esto significa que la representatividad de los resultados en cada una de las muestras respecto a la población general de trabajadores, no está garantizada debido al muestreo de conveniencia en ambos grupos nacionales. Debido a la diferente composición de las muestras de trabajadores nacionales, esto puede caracterizar la situación de comparar “manzanas con peras”, sin embargo, esta limitación es aparente porque los constructos evaluados pueden tener características

psicométricas estructurales similares o invariantes aún en la situación que difieran en la intensidad de estos constructos. Más aún, el nivel de invarianza alcanzado entre los grupos nacionales comparados (hasta la invarianza escalar, o invarianza fuerte; Meredith & Teresi, 2006), uno compuesto por dos grupos laborales diferentes (en la muestra peruana, educadores y trabajadores en general), y uno homogéneo, indica que la estructura interna de un instrumento como el PROPSIT puede ser invariante. Esto no señala que los resultados están completamente libres de restricciones para ser generalizados, pero respecto a la estructura del PROPSIT, hay un óptimo camino para avanzar en estudios interculturales más ambiciosos. Otros estudios de validación de medidas representativas de factores de riesgo psicosocial también han obtenido conclusiones importantes basadas en muestras heterogéneas de trabajadores (eg., Berthelsen et al., 2020; Hahr et al., 2014)

Se concluye que la estructura interna del PROPSIT entre los grupos de trabajadores peruanos y mexicanos, balanceados en las características del sexo, edad y grado de instrucción, es invariante respecto a la configuración de sus dimensiones (invarianza configuracional), la definición de cada dimensión (invarianza métrica) y la intensidad de respuesta de los ítems (invarianza de interceptos).

3.8. Objetivo 3d. Examinar la confiabilidad de los puntajes del PROPSIT

3.8.1. Introducción

Los puntajes observables de una medida, pueden derivarse luego de definir la estructura dimensional latente. Esta estructura dimensional con que el constructo es representado puede ser simple (eg., modelos unidimensionales) o compleja (eg., modelos multidimensionales), y habitualmente definen la manera cómo se va a obtener un puntaje observable, el mismo que se utiliza para describir a los evaluados, asociar variables o diferencia grupos. Por ejemplo, la autoeficacia ocupacional puede ser un constructo conceptualizado unidimensionalmente (eg., Rigotti, Schyns, & Mohr, 2008), y por lo tanto los ítems deben representar esta única dimensión identificada mediante estrategias de análisis factorial. Por el contrario, la autoeficacia ocupacional puede ser conceptualizado multidimensionalmente, con constructos de efectividad, perseverancia y adaptabilidad (Chaudhary, 2014) y, en consecuencia, se espera que sea operacionalizado con una medida multidimensional.

En una medida unidimensional o multidimensional, sus puntajes observados se usan para la práctica profesional y la evaluación, en lugar de puntajes derivados directamente de las técnicas de modelamiento (eg, puntajes factoriales obtenidos mediante técnicas de análisis factorial, o puntaje de atributo latente obtenidos del modelamiento de teoría de respuesta al ítem), posiblemente debido a que a) son más fáciles de interpretar, b) los puntajes de variables latentes son más difíciles de estimar porque implica que el usuario conozca las opciones diferentes opciones de modelamiento (eg, basado en la teoría clásica de los tests o en la teoría de respuesta al ítem; o basado en variables continuas o variables categóricas), c) requiere manejar eficazmente el software disponible, y d) el usuario requiere tener habilidades para tomar decisiones en su implementación. Por lo tanto, el puntaje observado directo es generalmente considerado como un proxy del atributo latente (Crişan, Tendeiro, & Meijer, 2020). Esto no es necesariamente un problema que signifique diferente estimación de la intensidad del evaluado en el atributo medido, porque en algunas teorías, el puntaje directo es una medida suficiente y

proxy del atributo latente, tanto en modelos paramétricos (Hemker, Sijtsma, Molenaar, & Junker, 1997; Sijtsma, & Hemker, 2000; van der Ark, 2005) como en modelos no paramétricos (eg., Mokken Scale Analysis; Crişan et al., 2020; van der Ark, 2005). Esto ocurre en la medida que se cumplen los supuestos de cada modelo.

Pero los puntajes observables continúan influenciados por la variabilidad aleatoria, es decir, el error de medición o el grado de variabilidad aleatoria que ocurre alrededor del puntaje observado, es importante para fines prácticos. De acuerdo a la *teoría clásica de los test* (TCT; Gulliksen, 1950, pp. 15), el error de medición puede reconocerse en la siguiente ecuación:

$$X = V + E \quad (1)$$

Donde X es el puntaje observado, V es el puntaje verdadero, y E es el error aleatorio. De acuerdo a la TCT, $\rho_{VE} = 0$, lo que significa que teóricamente V y E no están correlacionados. De acuerdo al AERA et al. (2014), la estimación de error de medición es un concepto vinculado con la confiabilidad de un puntaje, y es importante estimarlo para su uso aplicado. Desde la TCT, la confiabilidad (r_{xx}) es el grado de consistencia o reproducibilidad de respuesta (Brown, 2015; Gulliksen, 1950), y sirve para indicar la precisión de un puntaje. Mediante el cálculo de la confiabilidad, se puede estimar el grado de error o imprecisión, porque conceptualmente, $E = 1 - r_{xx}$ (E es el error estimado, y r_{xx} es una medida de precisión o confiabilidad; Gulliksen, 1950). Entonces, una estimación habitual del grado de precisión del puntaje es el error estándar de medición (Gulliksen, 1950):

$$S_E = S_X (1 - r_{xx}) \quad (2)$$

Donde S_E es el error estándar de medición, S_X es la desviación estándar del puntaje observado, y r_{xx} es la estimación de la confiabilidad (se han mantenido la simbolización en habla inglesa debido que forma parte de la literatura en teoría de medición).

La confiabilidad y/o precisión, y su medida derivada, el ESM, es una estimación única, pero puede no mantenerse constante entre grupos, o dentro de los niveles del puntaje (AERA et al., 2014). Efectivamente, como estimación sumaria, es útil como medida general de precisión, pero parece es difícil sostener que ésta pueda permanecer constante (i.e., igual) en diferentes condiciones y situaciones. Por ejemplo, una escala de estrés en la vida diaria, evaluada a un grupo de participantes en un contexto de moderada intensidad de estresores, probablemente arroje una confiabilidad más alta comparada con personas evaluadas con fuertes estresores, y donde la confiabilidad puede ser inferior (debido al estrecho rango de respuestas generadas por los participantes). De manera similar, los puntajes bajos de trabajadores en una medida de burnout pueden ser más precisos que los puntajes cercanos a la media. Esta variación de la precisión a lo largo de los niveles del atributo medido representa el error estándar de medición condicional (CSEM, por sus siglas en inglés; Meyer, 2014).

Hay unos pocos métodos para estimar el CSEM (eg., Meyer, 2014; Mollenkopf, 1949), y los resultados generalmente cambian de acuerdo al modelo teórico utilizado (Feldt, Steffen, & Gupta, 1985; Lahner et al., 2020). Pero un método que vincula el modelo de ecuaciones estructurales (SEM) con la teoría de respuesta al ítem (IRT) es la transformación de los parámetros estimados desde SEM hacia IRT (Kamata, & Bauer, 2008). Hasta la fecha, no parece haber algún estudio en la construcción y adaptación de medidas de factores psicosociales del trabajo, que introduzca esta metodología, y por lo tanto está ausente el conocimiento sobre la variabilidad del error de medición en diferentes niveles del puntaje, así como su confiabilidad condicional derivada de este enfoque (Raju, Price, Oshima, & Nering, 2007).

En este sentido, evaluar la confiabilidad de los puntajes del PROPSIT, y su potencial variabilidad en diferentes niveles del puntaje, es el objetivo del presente estudio. Este objetivo está subsumido en la evidencia de validez de estructura interna (APA, AERA, NCME, 2004).

3.8.2. Método

Participantes

Los participantes fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), luego de remover los participantes con potenciales respuestas con insuficiente esfuerzo. El tamaño de la muestra fue 305.

Instrumento

Escala de Procesos Psicosociales del Trabajo (Juárez-García, & Flores-Jiménez, 2020). El instrumento utilizado para el presente objetivo fue el obtenido del análisis de la dimensionalidad, y resuelto en el objetivo Objetivo 3a (Identificar la estructura dimensional del PROPSIT), mediante los métodos paramétricos. En los factores de riesgo psicosocial (PROPSIT – FRPS, 15 ítems), los factores establecidos fueron 4: demandas psicológicas (5 ítems), exigencias físicas (3 ítems), acoso psicológico (4 ítems), y demandas emocionales (3 ítems); en los factores recursos favorables (PROPSIT – FRF, 14 ítems), los factores establecidos fueron 3: recompensas y recursos (4 ítems), clima de apoyo social (5 ítems), y control laboral (5 ítems).

Procedimiento

Recolección de datos. Los participantes fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), y por lo tanto se aplica aquí los mismos procedimientos de recolección y conformidad con la conducta ética como fueron implementados.

Análisis. La estimación de la confiabilidad de consistencia interna se obtuvo con el coeficiente ω (McDonald, 1999) para cada puntaje del PROPSIT en las dos áreas, factores de riesgo psicosocial (FRPS) y factores recursos favorables (FRF). El coeficiente ω es una medida óptima de la confiabilidad de un puntaje cuando los ítems no tienen la misma validez o carga factorial (McDonald, 1999). Debido que los ítems fueron tratados como variables categóricas ordinales, ω fue obtenido con el método de Green y Yang (2009b), es decir, que se utilizó el estimador WLSMV (mean- and variance-adjusted

weighted least squares; Muthén, du Toit, & Spisic, 1997), y los umbrales correspondientes para cada una de las opciones de los ítems (ie., τ_i). Este método es comúnmente recomendado para el análisis factorial con variables categóricas (Brown, 2015),

Aunque el coeficiente α tiende a no ser recomendado (Green, & Yang, 2009a, 2009b), también fue calculado para propósitos de comparación, y porque su uso es predominante en la creación y adaptación de instrumentos de factores psicosociales del trabajo. Para ambos coeficientes, intervalos de confianza fueron generados por simulación bootstrap, como es recomendado por la literatura, debido que no hace presunciones sobre la forma de la distribución de los datos (Dixon, 2006). El método bootstrap fue el percentil, que es menos liberal (i.e., banda del intervalo ligeramente más amplias comparada con otros métodos bootstrap) en la condición de una medida de cinco ítems (Padilla, & Divers, 2013a, 2013b).

La diferencia entre los coeficientes α y ω se hizo mediante el método de intervalo de confianza para la diferencia $\alpha - \omega$ ($\Delta_{\omega - \alpha}$; Deng & Chan, 2017), el cual está basado en la estimación de la varianza tipo *sandwich* de las diferencias, y es de distribución libre (Deng & Chan, 2017). La confiabilidad condicional fue obtenida mediante un método que vincula las estimaciones del análisis factorial con el modelamiento IRT (Kamata, & Bauer, 2008). Este consistió en producir estimaciones equivalentes entre las cargas factoriales (ie., λ) y el parámetro IRT de discriminación (ie., parámetro a), mediante la fórmula:

$$a = \frac{\lambda}{\sqrt{1 - \lambda^2}} \quad (3)$$

El parámetro de dificultad del IRT (ie., δ_i) fue obtenido de los umbrales (ie., τ_i para cada punto de separación entre las opciones de respuestas) calculados en las correlaciones policóricas para cada opción de respuesta en los ítems:

$$\delta_i = \frac{\tau_i}{\sqrt{1 - \lambda^2}} \quad (4)$$

Para esta transformación, las cargas factoriales fueron obtenidas mediante un análisis factorial exploratorio con el estimador mínimo cuadrados no ponderados (ie., ULS), aplicada a las correlaciones policóricas entre los ítems. Los programas estadísticos utilizados fueron los paquetes *R psych* (Revelle, 2021), *MBESS* (Kelley, 2020) y *alphaomega* (Deng & Chan, 2017). El grado en que los coeficientes de confiabilidad fueron adecuados se hizo mediante de la *matriz de confiabilidad* (Ponterotto & Ruckdeschel, 2007), en que se toma en cuenta el tamaño de la muestra, el número de ítems y el valor numérico obtenido para definir a un nivel de confiabilidad en aceptable, moderado, bueno y excelente.

3.8.3. Resultados

Estimación de la confiabilidad

En la Tabla 37 se muestran los resultados de la confiabilidad de consistencia interna, estimados por el coeficiente α y ω . Se observa una clara tendencia de la relación entre el número de ítems y la magnitud de los coeficientes; la correlación Pearson entre ambos fue .914 y .938, para α y ω respectivamente. Las estimaciones puntuales de la confiabilidad en los puntajes de factores de riesgo psicosocial fueron alrededor de .80 en ambos coeficientes. Los coeficientes ω en las subescalas de factores psicosociales del trabajo ($M_\omega = .814$) y factores de recursos favorables ($M_\omega = .912$) tendieron a ser muy levemente superiores a los coeficientes α (respectivamente, .802 y .910). Se detectaron diferencias estadísticamente significativas en los puntajes de Demandas Psicológicas y Exigencias Físicas ($\Delta_{\omega-\alpha}$ fue levemente superior al valor 0.0), mientras que Demandas Emocionales y Acoso Psicológico, las diferencias no fueron estadísticamente significativas (Tabla 37). Respecto a los puntajes de Recursos favorables, las estimaciones α y ω variaron entre .868 y .945, y de acuerdo al intervalo de confianza de las diferencias ($\Delta_{\omega-\alpha}$) no hubo diferencias estadísticamente significativas entre ambos.

Tabla 37

Estimaciones de confiabilidad en el PROPSIT (n = 305)

	Nro ítems	ID Ítems	Estimaciones (ee)		$\Delta\omega - \alpha$ (ee)
			(IC 95%)		
			α	ω	
Riesgo Psicosocial					
FRPS 1	5	1, 2, 6, 10, 11	.865 (.01)	.872 (.01)	.005, .014
Demandas psicológicas			(.840, .882)	(.845, .895)	(.002)
FRPS 2	3	12, 13, 14	.788 (.02)	.785 (.02)	-.005, .016
Demandas emocionales			(.734, .837)	(.731, .839)	(.005)
FRPS 3	3	15, 16, 17	.793 (.02)	.806 (.02)	.010, .041
Exigencias físicas			(.727, .837)	(.772, .850)	(.008)
FRPS 4	4	18, 19, 20, 21	.818 (.04)	.839 (.02)	-.010, .052
Acoso psicológico			(.736, .900)	(.782, .896)	(.016)
Recursos Favorables					
FRF 1	4	24, 25, 35, 36	.868 (.01)	.877 (.01)	-.000, .003
Recompensas y recursos			(.837, .896)	(.848, .909)	(.000)
FRF 2	5	30, 31, 32, 33, 34	.946 (.006)	.945 (.007)	-.000, .001
Control laboral			(.929, .956)	(.928, .957)	(.000)
FRPS 3	5	37, 38, 39, 40, 41	.916 (.001)	.916 (.01)	-.000, .000
Clima de apoyo social			(.896, .936)	(.896, .935)	(.000)

Nota. IC generados por simulación Bootstrap, método percentilar (500 replicaciones). ee: error estándar. α : coeficiente de confiabilidad (modelo tau-equivalente). ω : coeficiente de confiabilidad (modelo congénico). $\Delta\omega - \alpha$: diferencia entre coeficientes α y ω .

Nivel de confiabilidad. De acuerdo a la matriz de confiabilidad (Ponterotto & Ruckdeschel, 2007), se nominó el nivel de adecuabilidad de los coeficientes como se presentan en la Tabla 38 . Se observa que la confiabilidad de los puntajes va desde el nivel aceptable hasta el nivel excelente; los puntajes de recursos favorables ocuparon predominantemente el nivel excelente, mientras que los puntajes de riesgos psicosociales consistentemente ocuparon los niveles moderado y bueno.

Tabla 38

Niveles de adecuación de los coeficientes de confiabilidad en los puntajes del PROPSIT

	Nro ítems	Aceptable	Moderado	Bueno	Excelente
Factores de Riesgo Psicosocial					
FRPS 1 Demandas psicológicas	5		X	X	
FRPS 2 Demandas emocionales	3		X	X	X
FRPS 3 Exigencias físicas	3	X	X	X	
FRPS 4 Acoso psicológico	4	X	X	X	
Factores de Recursos Favorables					
FRF 1 Recompensas y recursos	4			X	X
FRF 2 Control laboral	5				X
FRPS 3 Clima de apoyo social	5				X

Nota. N = 305. Zonas negras indican el rango de nivel de confiabilidad basado en los intervalos de confianza de los coeficientes de confiabilidad de cada factor.

Error estándar y confiabilidad condicional. Los resultados presentados en la Tabla 39 muestran la precisión, error estándar de medición condicional (CSEM), la información del puntaje (Test Info), y CSEM-estandarizado (ie, confiabilidad condicional basado en CSEM), a lo largo de varios niveles del atributo medido y en la métrica de puntajes estandarizados [-3.0, +3.0]. En la Figura 1 se puede ver con mayor claridad la presentación del CSEM y la confiabilidad condicional. En general se observa un parcial patrón regular en ambas áreas de evaluación (ie., riesgo psicosocial y recursos favorables): la precisión, y en consecuencia la confiabilidad, mejoran en una de las regiones extremas. Específicamente, en factores de riesgo psicosocial (excepto demandas psicológicas) la precisión mejora en la región de los puntajes más allá de una desviación estándar de la media; en contraste, en los factores de recursos favorables (excepto control laboral), la precisión y confiabilidad mejoran en los puntajes debajo de una desviación estándar de la media. Control laboral, presentó un patrón irregular (dos modas en ambas regiones extremas, debajo y encima de la media), mientras que demandas psicológicas presentó una forma regular pero centrada en el valor medio de su puntaje. Por último, en el área recursos favorables (Tabla 39), el nivel de la confiabilidad en el rango de los puntajes que va desde -3.0 hasta +3.0, tiende a ser alta debajo de una desviación estándar, mientras que, en el

área de riesgos psicosociales, la confiabilidad tiende a ser aceptable en el nivel ≥ 0.0 (es decir, sobre el puntaje promedio). La representación gráfica de la confiabilidad condicional puede ver en la

Figura 17.

3.8.4. Discusión

La versión del PROPSIT obtenida en el presente proyecto sugiere un desafío importante para la precisión de los puntajes, porque el predominante número de ítems en cada subescala (entre 3 y cinco) es menor que lo reportado en la práctica evaluativa. Esto es importante porque el número de ítems y el tamaño muestral están entre los factores que influyen la variabilidad del coeficiente α (Abdelmoula, Chakroun, & Akroun, 2015; İnal, Koğar, Demirdüzen, & Gelbal, 2017), y por lo tanto, uno de los motivos de las diferencias en el nivel de confiabilidad de los puntajes del PROPSIT proviene del tamaño de cada subescala. En general, medidas con un mayor número de ítems (> 3) tienden a ser diseñadas para presentar constructos unidimensionales o multidimensionales en la investigación en general (Meijer & Baneke, 2004; Rupp, 2013). En el contexto de las evaluaciones de factores psicosociales del trabajo, el número de ítems también tiende a ser mayor a 3 ítems. Por ejemplo, en la escala Effort–Reward Imbalance (ERI; Siegrist, Wege, Pühlhofer, & Wahrendorf, 2009), excepto la subescala esfuerzo (medido por 3 ítems), las subescalas de recompensa y sobre-compromiso tienen más de cinco ítems. Por otro lado, el Demand Control Support Questionnaire (DCSQ; Landsbergis, Theorell, Schwartz, Greiner, & Krause, 2000) con tiene tres subescalas con el número de ítems entre 5 y 6 (demandas psicológicas, control, apoyo social). Finalmente, el ISTAS-21 (Moncada, Llorens, Navarro, & Kristensen, 2005) contiene 3 subescalas entre 4 y 5 ítems, y excepcionalmente, dos escalas adicionales entre 3 y 2 ítems.

Tabla 39

Precisión y confiabilidad condicional para los puntajes del PROPSIT - FPST

	Niveles del atributo latente						
	-3	-2	-1	0	1	2	3
Factores de Riesgo Psic.							
FRPS 1							

Demandas psicológicas							
Test Info	3.21	4.76	6.25	6.89	5.93	3.75	3.17
CSEM	.56	.46	.40	.38	.41	.52	.56
Confiabilidad	.69	.79	.84	.85	.83	.73	.68
FRPS 2							
Demandas emocionales							
Test Info	.60	1.35	2.24	3.02	3.70	3.88	3.27
CSEM	1.30	.86	.67	.58	.52	.51	.55
Confiabilidad	.00	.26	.55	.67	.73	.74	.69
FRPS 3							
Exigencias físicas							
Test Info	.26	.71	1.93	3.58	3.12	4.53	4.54
CSEM	1.96	1.19	.72	.53	.47	.47	.47
Confiabilidad	.00	.00	.48	.72	.68	.78	.78
FRPS 4							
Acoso psicológico							
Test Info	.08	.28	.76	1.41	2.19	3.61	5.95
CSEM	3.47	1.90	1.15	.84	.68	.53	.41
Confiabilidad	.00	.00	.00	.29	.54	.72	.83

Niveles del atributo latente

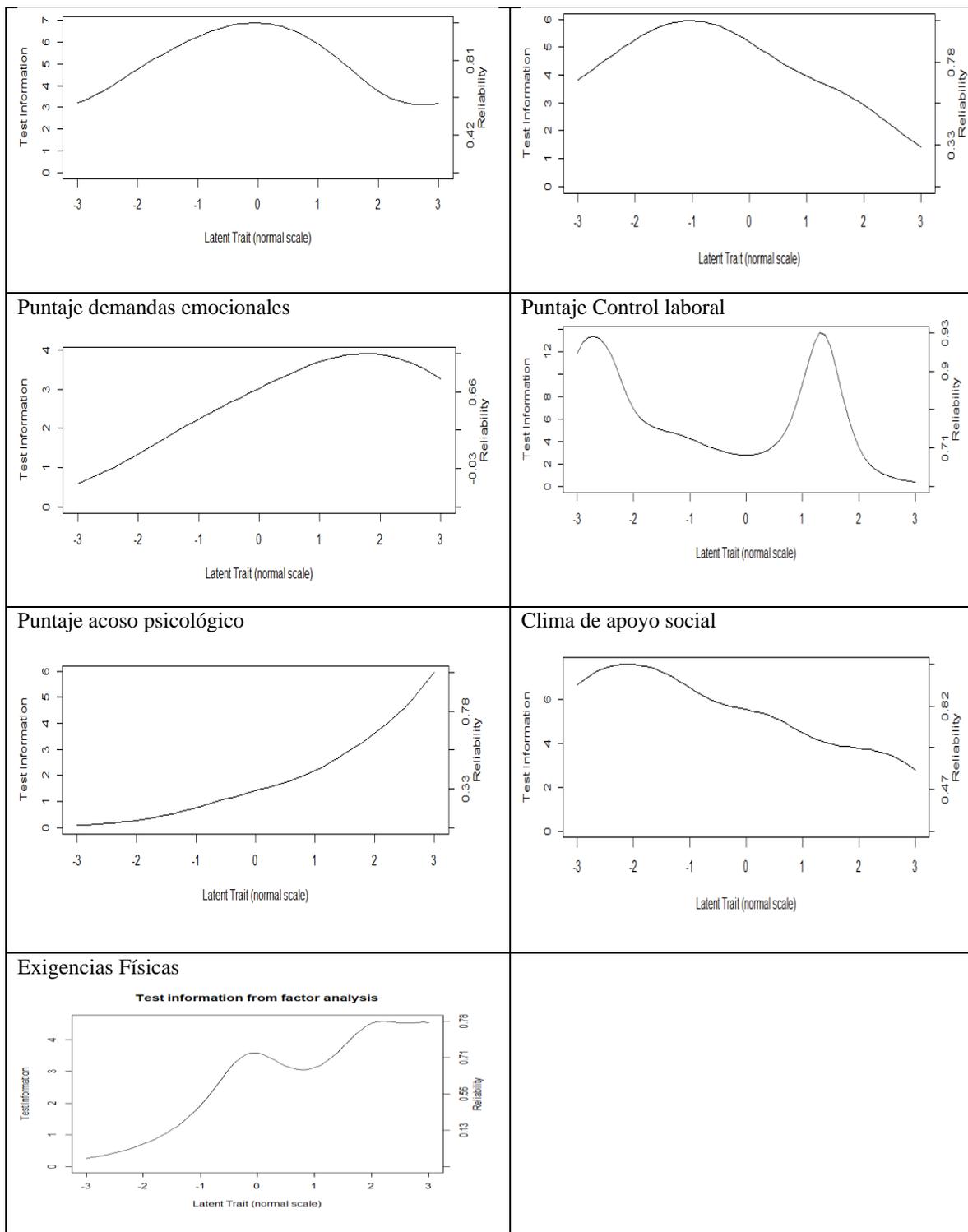
	-3	-2	-1	0	1	2	3
Factores Recursos Fav.							
RF1							
Recompensas y Recursos							
Test Info	3.82	5.27	5.95	5.19	3.95	2.92	1.41
CSEM	.51	.44	.41	.44	.50	.59	.84
Confiabilidad	.74	.81	.83	.81	.75	.66	.29
RF 2							
Control laboral							
Test Info	11.84	6.93	4.23	2.76	9.13	3.44	.36
CSEM	.29	.38	.49	.60	.33	.54	1.66
Confiabilidad	.92	.86	.76	.64	.89	.71	.0
RF 3							
Clima de apoyo social							
Test Info	6.66	7.59	6.54	5.54	4.49	3.78	2.79
CSEM	.39	.36	.39	.42	.47	.51	.60
Confiabilidad	.85	.87	.85	.82	.78	.74	.64

Nota. FPST: factores psicosociales del trabajo. Test Info: función de información del puntaje. SEM: error estándar de medición. Confiabilidad: SEM condicional estandarizado o equivalente a test info.

Figura 17

Función de información y confiabilidad condicional del PROPSIT - FPST

Factores de Riesgo Psicosocial	Factores de Recursos Favorables
Puntaje demandas psicológicas	Puntaje recompensas y recursos



Nota. La función de información del test fue derivada del análisis factorial (estimador: ULS).

Sin embargo, la práctica anecdótica y la literatura pueden indicar que como mínimo una escala debe contener 3 ítems (Marsh, Hau, Balla, & Grayson, 1998). Esta recomendación está centrada por motivos de metodología estadística, para evitar problemas de convergencia en las estimaciones

numéricas, y garantizar que los resultados sean válidos en contextos asintóticos, es decir, en muestras grandes (Mokkan & Lewis, 1982). Considerando que el PROPSIT-FPST es una medida para describir grupos de trabajadores en el contexto de evaluación masiva, y eficaz para el monitoreo, el número de ítems en las dimensiones del PROPSIT puede ser suficiente para conseguir una medición parsimoniosa, eficiente y válida de sus constructos medidos.

De acuerdo al método de intervalos de confianza de las diferencias ($\Delta\omega - \alpha$; Deng, & Chan, 2017), las diferencias estadísticamente significativas entre ambos coeficientes ocurrieron de manera infrecuente (solo en dos de las siete subescalas: demandas psicológicas y exigencias físicas). Pero dado que el límite inferior de este intervalo apenas superó el valor cero, podría concluirse que las diferencias son triviales, y para fines prácticos la estimación del coeficiente ω no es mejor que la confiabilidad expresada con el coeficiente α . Esta alta similaridad sugiere tres cosas: a) la confirmación de la unidimensionalidad latente de cada conjunto de ítems, b) la aparente propiedad de tau-equivalencia de los ítems (ie., ítems con igual carga factorial; Cronbach, 1951), y c) que los errores correlacionados son de tamaño trivial (alrededor de cero). Estos aspectos son características de las diferencias entre ambos coeficientes y el incumplimiento de algunas rompe la igualdad entre ambos coeficientes (Cronbach, 1951; Green, & Yang, 2009a). Una implicación práctica es que la confiabilidad de los puntajes de cada subescala del PROPSIT puede ser rutinariamente estimada con el coeficiente α , y que a menos que en futuras aplicaciones las diferencias en las cargas factoriales o el tamaño de errores correlacionados sean sustanciales, el coeficiente ω será una mejor estimación de la confiabilidad.

La adecuabilidad de los coeficientes (ie., el nivel cualitativo) en general no fue consistente, y ocupó entre una y tres categorías; esta amplitud estuvo influenciada parcialmente por el tamaño de la muestra (Abdelmoula et al., 2015), y en futuros estudios con el PROPSIT, un tamaño muestral más grande reducirá la amplitud del intervalo, y arrojará mayor precisión en la estimación. Casi todos los puntajes incluyeron el nivel bueno o más (excepto acoso psicológico y exigencias físicas), y dado la

brevidad de cada puntaje y el uso que se espera de ellos (eg., descripción de contextos psicosociales del trabajo, no individuos), estos niveles de confiabilidad pueden considerarse óptimos.

Respecto a la estimación condicional de la precisión y confiabilidad, excepto una subescala en cada una de las dos áreas del PROPSIT (ie, Factores de Riesgo Psicosocial y Factores de Recursos Favorables), se reconoció un patrón donde la precisión mejoró en uno u otro extremo del puntaje (ie., debajo o sobre la media). Aparentemente esto es un óptimo resultado porque en las regiones alejadas de la media pueden ser ubicados puntos de corte para separar grupos de alto y bajos puntajes, y son en estos grupos donde la clasificación debe ser más precisa. Específicamente, en el área de Factores de Riesgo Psicosocial, la precisión mejoró en los puntajes sobre la media, y esto es una característica óptima porque la intensidad fuerte con que los estresores ocurren en el ambiente laboral es un criterio para intervención. Por otro lado, en el área de Factores de Recursos Favorables, la ausencia de ambientes protectores, de apoyo social o de control en las tareas, es un criterio para intervención; y justamente, en los niveles debajo de la media de estos puntajes, la precisión de todas estas tendió a alta. También se observaron inconsistencias y las estimaciones subóptimas (por ejemplo, confiabilidad condicional de .00), pero pudieron ser influidas por el número de ítems, porque el CSEM también varía con el número de ítems (Lahner et al., 2020).

Entre las limitaciones del estudio, primero, no fueron estimados otros métodos de CSEM, y dado que las estimaciones de CSEM pueden cambiar sustancialmente de acuerdo al modelo utilizado (Lahner et al., 2020), se requiere extender las estimaciones de CSEM con otros modelos y evaluar sus discrepancias en los datos del PROPSIT. Entre los métodos basados en la TCT para CSEM, existe el método polinomial (Mollenkopf, 1949), elaborado a partir de la diferencia de las dos mitades del instrumento. Este método requiere el modelamiento de la diferencia entre mitades por medio de la regresión lineal múltiple polinomial, que incluye términos cuadráticos (x^2), cúbicos (x^3) y cuárticos (x^4) del puntaje. Otro método está basado en puntajes ajustados a una distribución beta-binomial (Meyer,

2014). Una segunda limitación fue la ausencia de comparaciones entre grupos sobre los coeficientes obtenidos. Los dos grupos muestreados (ie., trabajadores de varias profesiones y puestos, y educadores) puede mostrar diferentes patrones de confiabilidad y error de medición. Sin embargo, debido al tamaño de ambos grupos (< 200), probablemente ocurriría error Tipo II. Una investigación en que presente la evidencia comparativa entre grupos requerirá tamaños muestrales alrededor o superiores a 300 (Ponterotto, & Ruckdeschel, 2007) o 400 por grupo (Charter, 1999, 2001).

3.9. Objetivo 4. Examinar la relación de los constructos del PROPSIT – FPST con otras variables

2.9.1 Introducción

En la exploración de la red de efectos que pueden tener los factores psicosociales del trabajo en la salud de las personas, existe numerosa evidencia de la sintomatología que sirve como marcador del efecto de la exposición a factores tales como el desequilibrio esfuerzo/recompensa, el limitado apoyo percibido, las altas demandas laborales, demandas emocionales y mobbing (o ambiente hostigador), etc. La ansiedad y depresión se encuentran entre este tipo de sintomatología (Magnavita, & Fileni, 2014). Los síntomas de ansiedad y depresión en general son marcadores generales de la salud mental y el bienestar (Wijk, 2021), y como las áreas de actividad laboral están subsumidas en este contexto más general de actividad humana, ambos síntomas (eg., ansiedad y estrés) también pueden ser marcadores de efectos emocionales de exposiciones a ambientes laborales altamente estresantes. La ansiedad y estrés en general son considerados piezas básicas o signos emocionales vitales del estado emocional de las personas, según investigaciones con pacientes clínicos y con aquellos provenientes de la comunidad en general (Spielberger, & Reheiser, 2009). Efectivamente, respecto al área de estudio de los factores psicosociales del trabajo (FPST), los síntomas de ansiedad (Andrea, Bültmann, van Amelsvoort, & Kant, 2009; Duru, Ocaktan, Çelen, & Örsal, 2018; Magnavita, & Fileni, 2014) y de depresión (Balducci, Avanzi, & Fraccaroli, 2014; Duru et al., 2018; Berthelsen et al., 2015; Magnavita, & Fileni, A. 2014; Theorell et al., 2015; Żołnierczyk-Zreda, & Holas, 2018;), han sido asociados al desequilibrio negativo del esfuerzo y recompensa, al pobre apoyo social en el trabajo, a las altas demandas laborales y emocionales, y la intensidad del ambiente hostigador (también conocido como mobbing).

Por otro lado, existen procesos cognitivos que reconociblemente se deterioran por la exposición a factores de riesgo psicosocial, y que directamente se asocian al pobre desempeño en la tarea. En general, este deterioro puede caracterizarse en aspectos como la toma de decisiones, el recordar hechos,

pensar claramente y la atención focalizada (Stenfors, Magnusson Hanson, Oxenstierna, Theorell, & Nilsson, 2013). Entre estas quejas cognitivas auto-informadas asociadas con diferentes aspectos psicosociales del trabajo, hay una asociación negativa con la memoria de trabajo (Boschi, Trenoweth, & Sheppard, 2017; Eskildsen, Andersen, Pedersen, Vandborg, & Andersen, 2015; Horvat, & Tement, 2020), particularmente en condiciones de estrés prolongado (Luethi, Meier, & Sandi, 2008). De hecho, esto supone que la intensidad de este problema interacciona con condiciones psicosociales laborales (Stenfors et al., 2013).

Los síntomas emocionales y de conducta desadaptativa asociados a los factores de riesgo psicosocial (FRPS) no solo se restringen a sus expresiones particulares o conjuntos sintomáticos de ansiedad y depresión, sino que también son expresados en la percepción general de salud mental y con medidas genéricas (Duru et al., 2018; Harvey et al., 2017; Nieuwenhuijsen, Bruinvels, & Frings-Dresen, 2010; Stansfeld & Candy, 2006). Esto es posible debido que los contenidos específicos de muchas medidas generales de salud mental incluyen conductas representativas de síntomas específicos, como problemas para dormir, percepción de valía personal, entre otros. Por lo tanto, la percepción de deterioro de la salud mental es una respuesta sensible a la intensidad de los estresores generales, y en particular a los estresores laborales.

Por otro lado, ambos conjuntos dimensionales (factores de riesgo psicosocial y factores de recursos favorables) no solo pueden entenderse por sus efectos negativos incrementados o decrementados, sino también por sus efectos positivos (o la disminución de los mismos). Por ejemplo, el entusiasmo laboral (o work engagement) puede ser entendido dentro de un marco de constructos motivacionales desarrollado en el contexto laboral (Inceoglu, & Fleck, 2010), y que dentro del modelo demandas-recursos laborales (DRL; Bakker, & Demerouti, 2007), forma parte del proceso motivacional que involucra los resultados positivos de la actividad laboral (e.g., satisfacción, eficiencia, etc.) y los recursos laborales y personales (Schaufeli, 2013). Esto tiene una clara implicación para los

estudios de validación de medidas de factores psicosociales, porque debido que los recursos favorables son aspectos antecedentes de la satisfacción o eficiencia laboral (Schaufeli, 2013), los puntajes de medidas de engagement deben mantener dependencia monotónica con tales aspectos.

Basado en los estudios previos sobre el amplio rango de respuestas conductuales, emocionales y de salud vinculados con lo FPST, una medida que evalúe estas condiciones del ambiente laboral debe probar su relación con estas respuestas esperadas. El análisis de las asociaciones de las dimensiones de una medida de FPST es la evidencia de validez requerida en la descripción y clasificación de ambientes laborales y en el conocimiento de sus relaciones teóricas (AERA et al., 2014). Por lo tanto, el objetivo del presente estudio es examinar la relación entre las dimensiones del Cuestionario de Evaluación de Procesos Psicosociales del Trabajo (PROPSIT; Juárez-García, & Flores-Jiménez, 2020) y medidas de efecto emocional, laboral y conductual. EL PROPSIT fue creado para representar los factores del contexto laboral, moderadores, efectos y variables psicosociales relevantes en trabajadores mexicanos con contrato formal. Es una medida de autoinforme que, aún con las limitaciones típicas de este tipo de medidas, está alineada con la práctica habitual sobre el uso de medidas en auto-informe en la investigación de FPST (Magnavita, 2008). El modelo-marco de esta medida fue demandas-recursos laborales (Bakker, & Demerouti, 2007), y desde el cual se conceptualizaron las áreas de factores, procesos y efectos. El PROPSIT fue construido con una visión relevante para su contexto mexicano de origen, pero potencialmente generalizable a otros contextos latinoamericanos, debido que sus dimensiones teóricas tienen aparente universalidad étic (i.e., factores consistentemente identificados en varios estudios internacionales; Broetje, Jenny, & Bauer, 2020), y que son consistentemente elegidos como atributos centrales para evaluar factores de riesgo psicosocial o factores psicosociales favorables (recursos).

Aunque la dependencia entre las variables durante la construcción y adaptación de medidas se hace con un enfoque de relaciones lineales en particular, y en las ciencias sociales en general (Speed,

2011), es razonable que existan otros tipos de asociación. En la detección de asociaciones bivariadas, pueden existir asociaciones lineales, monotónicas y no monotónicas (Fujita et al., 2009), y la investigación psicológica no ha permanecido ajeno a la exploración de asociaciones diferentes a la dependencia lineal entre variables. Efectivamente, en la literatura psicológica de la investigación de constructos del desarrollo y en otros contextos aplicados, las relaciones no lineales han aparecido como un modelo de covariabilidad para representar válidamente la asociación entre las variables (Makonnen, 2019). Por ejemplo, en el desarrollo de habilidades visomotoras en un rango de edad sensible de cambio (e.g., Merino, 2011), en el cambio de roles, estatus y personalidad en adultos (e.g., Helson, & Soto, 2005), y en la expresión de sintomatología depresiva (e.g., Besser, Priel, Flett, & Wiznitzer, 2007). En el contexto laboral, también se han explorado no-linealmente la relación de las habilidades de afrontamiento y estrés en adultos trabajadores (e.g., García-Arroyo, & Osca, 2017), y entre el desempeño laboral y la inseguridad laboral y afectividad negativa (Mäder, & Niessen, 2017). Seguramente existen otros muchos ejemplos de aplicación de modelos no lineales, pero en general la identificación de asociaciones monotónicas no lineales y asociaciones no-monotónicas no parece ser frecuente. Especialmente útiles en contextos exploratorios, la evaluación de la dependencia estadística entre las variables requiere detectar de algún forma de relación funcional bivariada (Clark, 2013; Reshef et al., 2011), sin necesariamente iniciar la exploración desde una hipótesis definida sobre algún tipo de asociación, como ocurre habitualmente cuando se inicia con la búsqueda de relaciones lineales (Clark, 2013; de Siqueira Santos, Takahashi, Nakata, & Fujita, 2014; Dümcke, Mansmann, & Tresch, 2014; Fujita et al., 2009). Esto sugiere que la búsqueda de relaciones bivariadas se inicie con la detección de algún tipo de asociación existente, y que sirva como condición para realizar un segundo paso donde se prueben relaciones bivariadas específicas, como la asociación de tipo lineal. En este sentido, la incorporación de métodos modernos para la detección de un amplio rango de asociaciones

bivariadas en ciencias de la conducta son oportunidades aplicables también en la adaptación de medidas psicosociales del ambiente laboral.

Se plantearon las siguientes hipótesis.

H₁: los factores de riesgo psicosocial mostrarán dependencia monotónica incremental con efectos emocionales (experiencia de estrés y síntomas depresivos), y laborales (accidentes laborales menores y dificultades cognitivas), y efecto decremental con efectos motivacionales (entusiasmo laboral).

H₂: los factores de recursos favorables mostrarán dependencia monotónica decremental con efectos emocionales (experiencia de estrés y síntomas depresivos), y laborales (accidentes laborales menores y dificultades cognitivas), y efecto incremental con efectos motivacionales (entusiasmo laboral).

2.9.2 Método

Participantes

Los participantes fueron los mismos del Objetivo 3a (Identificar la estructura dimensional del PROPSIT), luego de remover los participantes con potenciales respuestas con insuficiente esfuerzo. El tamaño de la muestra fue 305 participantes. La distribución de valores perdidos relevantes a las variables medidas aparece en la Tabla 40.

Tabla 40
Distribución de valores perdidos

	Trabajadores en general (n = 184)	Educadores (n = 121)	Total (n = 305)
UWES	Missing = 0	No recolectados	Missing = 121
SOS	Missing = 0	Missing = 25	Missing = 25
PHQ-4	Missing = 0	Missing = 9	Missing = 9
Stress	Missing = 0	Missing = 0	Missing = 0
Accidentes	Missing = 0	Missing = 0	Missing = 0
Dificultades cognitivas	Missing = 0	Missing = 0	Missing = 0

Nota. Missing: número de respuestas perdidas.

Instrumentos

Escala de Procesos Psicosociales del Trabajo (Juárez-García, & Flores-Jiménez, 2020). El instrumento utilizado para el presente objetivo fue el obtenido del análisis de la dimensionalidad, y resuelto en el objetivo Objetivo 3a (Identificar la estructura dimensional del PROPSIT), mediante los métodos paramétricos. En los factores de riesgo psicosocial (FRPS), los factores establecidos fueron 4: demandas psicológicas (5 ítems), exigencias físicas (3 ítems), acoso psicológico (4 ítems), y demandas emocionales (3 ítems); en los factores de recursos favorables (FRF), los factores establecidos fueron 3: recompensas y recursos (4 ítems), clima de apoyo social (5 ítems), y control laboral (5 ítems).

Escala de Sobrecarga de Estrés - Breve (SOS-S; Amirkhan, 2018). Medida multidimensional de auto-informe que evalúa la percepción de sobrecarga debida al estrés (10 ítems), mediante dos dimensiones: vulnerabilidad personal (5 ítems) y carga de eventos (5 ítems). Los ítems están escalados ordinalmente, desde “nada” hasta “mucho”. Para el presente estudio, los ítems fueron traducidos por un método de consenso en un panel de 5 investigadores latinos, psicólogos, con dominio del idioma inglés (entre 10 y 20 años de experiencia de trabajo con literatura científica en inglés). En el presente estudio, se utilizó un puntaje único debido a la alta correlación entre las dimensiones ($r = .74, p < .01$), algo consistente con otros estudios donde ocurrió solapamiento expresado en altas correlaciones entre las dimensiones, y numerosos ítems con cargas cruzadas (Duan, & Mu, 2018; Van Wijk, 2021; Wilson et al., 2018). La confiabilidad del puntaje total fue $\omega = .86$, bootstrap-IC 95% = .82, .89).

Ítem Único de Estrés (Elo, Leppänen & Jahkola, 2003). Está destinado para evaluar la intensidad del estrés en un contexto general o de la vida diaria. El respaldo para su validez tiende a ser favorable (Elo et al., 2003; Littman, White, Satia, Bowen & Kristal, 2006; Salminen, Kouvonen,

Koskinen, Joensuu & Väänänen, 2014). En general, es medida eficiente para aproximarse a la experiencia del estrés, y en particular ha sido utilizado en estudios sobre estrés laboral (Arapovic-Johansson, Wåhlin, Kwak, Björklund, & Jensen, 2017; Elo et al., 2003; Houdmont et al., 2019; Omholt, Tveito, & Ihlebæk, 2017; Salminen et al., 2014).

Escala de Engagement de Utrech (UWES-3; Schaufeli, Shimazu, Hakanen, Salanova, & De Witte, 2019). Evalúa el entusiasmo laboral, con tres ítems sobre vigor, dedicación y absorción elegidos con base racional y empírica por los autores y derivados del UWES-9. El escalamiento de respuesta es de siete puntos, entre 0 (*nunca*) y 6 (*siempre*). Evidencias de validez de constructo favorables han sido reportado en Merino-Soto et al., (2022). En el presente estudio, la estructura interna fue adecuada, $\omega = .94$ (IC 95% = .91, .97).

Patient Health Questionnaire-4 (PHQ-4; Kroenke, Spitzer, Williams, & Löwe, 2009).

Medida breve de despistaje de síntomas emocionales y cognitivos de depresión (dos ítems) y ansiedad generalizada (ítems); es internacionalmente aceptada como una medida de despistaje total de eficiente distrés psicológico. Está escalada con cinco puntos (desde *para nada* hasta *casi todos los días*) en un marco de las últimas dos semanas. En estudios con participantes peruanos (Merino-Soto, Dominguez-Lara, & Fernández-Arata, 2017; Merino-Soto, Angulo-Ramos, & López-Fernández, 2019), sus ítems de ansiedad y depresión correlaciones mantienen coherencia teórica con otros constructos. En los presentes datos, el ajuste a dos dimensiones correlacionadas (i.e., ansiedad y depresión) fue satisfactorio, $WLSMV-\chi^2 = .120$ (gl = 1, $p = 1.0$) CFI = 1.00, SRMR = .00, WRMR = .061, pero la correlación inter-factorial fue muy alta ($r = .xx$, IC 95% = .88, .99); por lo tanto, se utilizó un puntaje único ($WLSMV-\chi^2 = 3.01$, gl = 2, $p = .22$, CFI = 1.00, SRMR = .02, WRMR = .409; cargas factoriales: .89, .90, .95 y .93), interpretado como distrés psicológico (Kroenke et al., 2009); la consistencia interna del puntaje fue $\alpha = .90$ (IC 95% = .85, .93).

Dificultades cognitivas. Se creó una medida de ítem único que sirvió como *un proxy* de problemas globales de naturaleza cognitiva en el contexto laboral, asociados con la atención, memoria y desempeño preciso, que son aspectos sensibles al estrés (Setterlind, & Larsson, 2015; Boschi et al., 2017; Horvat, & Tement, 2020; Eskildsen et al., 2015). El contenido fue: ¿Con qué frecuencia ha tenido problemas de memoria (por ejemplo, olvidar dónde puso las cosas), atención (por ejemplo, pobre concentración) o acción (por ejemplo, hacer mal algo) en el trabajo?, con cinco opciones ordinales (nada, raramente, ocasionalmente, con frecuencia, muy frecuentemente).

Accidentes menores en el trabajo (Juárez-García, 2007). Se utilizó una medida de autoinforme de lesiones laborales de un solo ítem, para representar una aproximación a la frecuencia de los incidentes menores ("En los últimos 12 meses, ¿ha tenido accidentes menores en el trabajo [cortes o contusiones leves] que no requirieron atención médica?"). Las opciones de respuesta fueron: ninguno, rara vez, ocasionalmente, frecuentemente, muy frecuentemente. Debido a su relación con la percepción de un entorno seguro (Abbas, Zalat, & Ghareeb, 2013; Nielsen & Mikkelsen, 2007), y a su eficacia para identificar áreas clave de accidentes, obtener prevalencias, monitorizar la seguridad laboral y contrastar con la información proporcionada por los Empresarios, este tipo de medidas de un solo ítem suelen aplicarse en encuestas epidemiológicas nacionales o en estudios independientes (por ejemplo, Free, Groenewold, & Luckhaupt, 2020; Jazari et al, 2018; Leibler & Perry, 2017; Marcum, Chin, Anderson, & Bonauto, 2017).

Procedimiento

Recolección de datos. Los participantes fueron los mismos del Objetivo 4a (Identificar la estructura dimensional del PROPSIT), y por lo tanto se aplica aquí los mismos procedimientos de recolección y conformidad con la conducta ética como fueron implementados.

Análisis. El análisis de la asociación entre las dimensiones del PROPSIT en las áreas de factores de riesgo psicosocial (FRPS: demandas psicológicas, demandas emocionales, exigencias físicas y acoso psicológico) y factores de recursos favorables (RFR: recompensas y recursos, control laboral y clima de apoyo social), con las variables externas (UWES, SOSS, PHQ-4, accidentes laborales menores y dificultades cognitivas) fue realizado con un enfoque bivariado. El enfoque bivariado permitió presentar un número único que resumen el grado de asociación, usual en la investigación de la cuantificación de los estudios de validez de constructo (Westen & Rosenthal, 2003).

La evaluación de las relaciones bivariadas entre las dimensiones del PROPSIT y las variables externas consistió en aplicar medidas de dependencia entre variables aleatorias. Esta fue implementada de la siguiente manera: primero, se implementaron gráficos de dispersión entre las variables asociadas, para identificar visualmente la forma de la relación; esto es, ausencia de aparente asociación, asociación no lineal, asociación monotónica, o asociación lineal. El gráfico utilizado fue de ajuste de regresión polinomial local, con grado 2 y tipo gaussiano, con el que se obtiene un gráfico de dispersión conteniendo una línea de ajuste suavizada y consistente con la variación local (esto es, aplicados a una región bivariada de las variables correlacionadas) de los datos (Cleveland, Grosse, & Shyu, 1992).

Una vez identificada visualmente la asociación, se implementó una medida de dependencia monotónica. La medida dependencia monotónica fue el coeficiente de correlación *rho* de rangos de Spearman (1904); *rho* es menos afectado por valores extremos comparado con otras medidas de asociación no lineal (Alsayed, & Manzi, 2019). Dado que es recomendable explorar las asociaciones bivariadas con diferentes métodos lineales y no lineales (Wang et al., 2015), finalmente, dado que el coeficiente de producto de momentos de Pearson es el método estándar para establecer el tamaño del efecto de asociaciones (Funder & Ozer (2019), también fue implementado. Las asociaciones aplicadas a las medidas de ítem único (accidentes menores en el trabajo y dificultades cognitivas) fueron analizadas también con el coeficiente de Pearson, dado que las mediciones ordinales pueden ser tratada

como variable continua cuando el número de categorías ordinales sugeridas es 5 o más 6 (Khamis, 2008; Robitzsch, 2020). La interpretación de la magnitud de las correlaciones fue establecida con la propuesta de Funder y Ozer (2019), para los niveles muy pequeño ($r = .05$), pequeño ($r = .10$), moderado ($r = .20$), grande ($r = .30$), y muy grande ($r = .40$).

Los análisis fueron implementados con los programas R *DescTools* (Signorell et al., 2021), *bootcorci* (Rousselet, Pernet, & Wilcox, 2019) y *stats* (R Core Team).

2.9.3 Resultados

Visualización de las asociaciones bivariadas. La asociación bivariada representada en las figuras Figura 18, Figura 19, Figura 20, Figura 21, Figura 22, y Figura 23 (ver estas figuras al final del texto), mostraron constante variación en la forma de cada curva graficada. En general, las asociaciones representadas tendieron a no ser lineales, e incluyeron alguna heterogeneidad local. Esta heterogeneidad fue presentada por curvaturas en que la variable externa cambió de patrón asociativo, pero fueron de aparente cambio pequeño. Por ejemplo, en la Figura 20 (estrés – ítem único con demandas psicológicas y clima da de apoyo). En todas las figuras hubo aparente fuerte dispersión de los datos, sugiriendo que las correlaciones estuvieron generalmente limitadas a magnitudes moderadas o moderadamente altas.

Asociaciones bivariadas

Factores de Riesgo Psicosocial. Con la sintomatología de distrés (PHQ-4), todos los puntajes de Factores de Riesgo Psicosocial fueron estadísticamente significativos, y en la dirección hipotetizada (signo positivo); en el tamaño de estas correlaciones fue pequeño (Tabla 41). La excepción fue Demandas psicológicas (FRPS1), asociado moderada y negativamente con PHQ-4. Con la medida de sobrecarga de estrés (SOS-S), las correlaciones fueron variables, porque estuvieron desde el nivel pequeño (FRPS3) hasta grande (FRPS1), pero esencialmente en la dirección hipotetizada. Las correlaciones con la medida única de estrés fueron también pequeñas (alrededor de .05) y en la

dirección hipotetizada (signo positivo). Con el puntaje de engagement (UWES), las correlaciones fueron negativas, y de tamaño moderado ($> |.20|$) o grande ($> |.30|$). Las demandas psicológicas de las tareas laborales (FRPS1) fueron de diferente orientación a la hipotetizada y comparativamente mayores que el resto de los otros factores; específicamente, FRPS1 consistentemente mostró correlaciones monotónicas en dirección negativa con PHQ4, SOS-S, accidentes menores y dificultades cognitivas, y correlación positiva con UWES. Demandas físicas mostró una correlación marcadamente diferente (y correlación muy alta, $> .35$) con accidentes menores en el trabajo, respecto al resto de las dimensiones (Tabla 41).

Factores de Recursos Favorables. Con el PHQ-4, las correlaciones obtenidas tendieron a ser moderadamente o altas, con SOS-S fueron altas, con accidentes menores fueron bajas, y con dificultades cognitivas fueron bajas o altas (Tabla 41, encabezado Factores de Recursos Favorables). Todas las dimensiones mostraron un consistente teórico y cuantitativo (i.e., el patrón de correlaciones negativas y positivas). Particularmente, recompensas y recursos mostró correlación alta ($> .30$) con dificultades cognitivas y distrés psicológico, además de ser muy diferente del resto de las dimensiones. La asociación con el ítem único de estrés general no fue estadísticamente significativa con ninguna dimensión, y, por el contrario, todas las dimensiones correlacionaron en magnitud muy alta con el UWES.

Discusión

Para alcanzar este objetivo, el análisis de las asociaciones bivariadas fue hecha por dos medios: gráfica y cuantitativa. En la estimación gráfica, la línea suavizada obtenida indicó la existencia de asociaciones ligeramente curvilíneas, un tipo de asociación que no puede ser capturas apropiadamente por las correlaciones lineales o monotónicas. También se hallaron relaciones monotónicas y aproximadamente lineales, lo que sugiere que la aplicación de métodos de asociación monotónica o lineales pueden ser recomendadas. Las relaciones observadas en las figuras indicaron que su magnitud

difícilmente puede ser muy alta o alta, debido que la dispersión de los puntos bivariados alrededor de las líneas suavizadas *loess* fueron amplias. Una implicación de esto es que los resultados no fueron subestimados por el tipo de coeficiente aplicado (i.e., coeficiente de Spearman), sino que la tendencia en las asociaciones no fue bien definida como para alcanzar magnitudes altas.

Tabla 41

Asociación monótonica entre el PROPSIT y variables externas

	PHQ-4	SOS-S	Estrés	Accidentes menores	Dificultades cognitivas	UWES
Factores de Riesgo Psicosocial						
FRPS 1 Demandas psicológicas	-.282* (-.38, -.17)	-.299* (-.41, -.19)	.202* (.09, .309)	-.118* (-.22, -.01)	-.267* (-.36, -.15)	.459* (.32, .57)
FRPS 2 Demandas emocionales	.198* (.09, .31)	.209* (.09, .31)	.081 (-.03, .19)	.100 (-.00, .20)	.187* (.06, .28)	-.230* (-.36, -.08)
FRPS 3 Exigencias físicas	.148* (.03, .25)	.126* (.00, .25)	.074 (-.03, .18)	.396* (.28, .48)	.109 (-.00, .20)	-.238* (-.38, -.10)
FRPS 4 Acoso psicológico	.143* (.02, .026)	.198* (.07, .31)	.103 (-.00, .21)	.177* (.06, .28)	.024 (-.08, .13)	-.318* (-.45, -.18)
Factores de Recursos Favorables						
FRF 1 Recompensas y recursos	-.442* (-.52, -.34)	-.438* (-.54, -.32)	.014 (-.10, .14)	-.188* (-.29, -.07)	-.360* (-.46, -.25)	.729* (.65, .79)
FRF 2 Control laboral	-.240* (-.35, -.11)	-.354* (-.46, -.22)	.027 (-.08, .15)	-.146* (-.24, -.03)	-.157* (-.27, -.03)	.733* (.64, .80)
FRF 3 Clima de apoyo	-.266* (-.37, -.15)	-.293* (-.40, -.17)	.009 (-.11, .11)	-.155* (-.26, -.04)	-.125* (-.23, -.02)	.560* (.43, .65)

Nota. PHQ-4: medida de distrés psicológico. SOS-S: medida de sobrecarga del estrés. Estrés: ítem único de estrés. Accidentes menores: ítem único de accidentes menores en el trabajo. Dificultades cognitivas: ítem único de dificultades de memoria y atención. UWES: Utrecht work engagement scale. Intervalos de confianza en 95%, generados por simulación bootstrap (n = 500). *p < .05

De acuerdo los resultados, la hipótesis de validez planteada para los factores de riesgo psicosocial es parcialmente aceptada, debido que, mientras las dimensiones presentaron predominante relación positiva (por ejemplo, como teóricamente puede esperarse con sobrecarga de estrés[SOS-S]), FRPS1 (demandas psicológicas vinculadas con la tarea) consistentemente mostró correlaciones en

dirección negativa con PHQ4, SOS-S, accidentes menores y dificultades cognitivas, y correlación positiva con UWES. Estos resultados sintetizan y completan lo hallado en el nivel de los ítems (Objetivo 1c: estructura de respuesta y validez con criterios externos), y en conjunto conducen a interpretar que, la existencia de ambientes con alta exigencia psicológicas para cumplir con el desempeño laboral se asocia con una disminución del estrés psicológico, sobrecarga de estrés, accidentes y dificultades cognitivas, mientras que crea un incremento con el engagement. Esto es contraintuitivo porque en general, altas demandas de la tarea están asociadas con un incremento en estas variables externas; sin embargo, esto puede ser un efecto de la baja intensidad de las demandas registradas en el grupo de participantes. Esto es, que los participantes no reportaron demandas frecuentemente muy intensas, y en consecuencia restringieron la mayor densidad de respuesta en las zonas moderadas o bajas en el constructo. Estos niveles pueden suficientemente energizantes para dedicarse a las tareas laborales de manera eficiente, un aspecto coherente con el modelo demandas/recursos (Bakker, & Demerouti, 2014) cuando la interacción entre los recursos y demandas es efectiva y positiva para el trabajador, mediante la amplificación de los recursos. Por otro lado, la débil naturaleza curvilínea de la relación de algunas variables con la medida de estrés SOS-S, también indica una tendencia que posiblemente el coeficiente de Spearman (i.e., de dependencia monótonica) no capturó completamente. Algunas asociaciones particulares deben ser resaltadas: la medida de accidentes menores en el trabajo se asoció muy alto con la dimensión de exigencias físicas, una coherente covariación entre ellas debido que la intensidad de la actividad física implicada en la tarea laboral puede aumentar la exposición a accidentes. Y demandas psicológicas fue comparativamente alto con dificultades cognitivas, una covariación razonable debido a las respuestas cognitivas implicadas del trabajador.

Respecto a la hipótesis de validez para las dimensiones de recursos favorables, esta es aceptada, porque la dirección de la asociación monótonica fue consistente en todos los coeficientes obtenidos, y

también fue consistente con la literatura. Particularmente, el control laboral mostró mayor asociación con la sobrecarga de estrés (ie, puntajes del SOS-S), y la percepción de control (y la ausencia del mismo) es definicional para la experiencia del estrés (Amirkhan, 2018), y el contenido parcial del SOS-S contiene ítems referidos a la pérdida de control.

De acuerdo a Funder y Ozer (2019), las estimaciones pequeñas son razonables y creíbles debido a la larga evidencia de correlaciones predominantemente de similar tamaño en la investigación en ciencias conductuales. Por lo tanto, las estimaciones de tamaño pequeño (obtenidas aquí en las escalas de factores de riesgo, pero en las escalas de recursos), no deben sorprender o conducir a pensar en la influencia de fuertes sesgos. Las estimaciones de correlaciones obtenidas en este estudio están dentro del tipo de estimaciones que pueden considerarse *confiables* y potencialmente replicables (Funder, & Ozer, 2019), debido a algunas condiciones óptimas: el instrumento principal (ie., PROPSIT) está configurado claramente en distintas dimensiones con moderada o pequeña dependencia entre ellas, las confiabilidades de los puntajes son altas, y las medidas externas son unidimensionales y con confiabilidades altas.

Los resultados respecto a los síntomas de depresión y ansiedad han estado en acuerdo con la tendencia hallada de su asociación con los factores de riesgo psicosocial (Andrea et al., 2009; Balducci et al., 2014; Berthelsen et al., 2015; Duru, et al., 2018; Magnavita, & Fileni, 2014; Theorell et al., 2015; Żołnierczyk-Zreda, & Holas, 2018), y así como con las respuestas de estrés (Amirkhan, 2018; Arapovic-Johansson et al., 2017; Elo et al., 2003; Houdmont et al., 2019; Omholt et al., 2017; Salminen et al., 2014). También, la investigación sobre el UWES habitualmente indica una relación de tipo incremental con los factores vinculados a los recursos laborales (eg, recompensa, apoyo social laboral), y de tipo decremental con los factores definicionales del riesgo psicosocial laboral, como fuertes demandas psicológicas. De acuerdo a los resultados obtenidos en el presente estudio, se concluye que las dimensiones del PROPSIT siguen esta línea de antecedentes, por lo tanto, la validez de las

dimensiones respecto a la red de constructos y criterios laborales indica la apropiada construcción de su adaptación, así como la interpretación de sus puntajes.

Entre las limitaciones de estudio, primero, la representatividad de la muestra no está asegurada debido a su naturaleza no probabilística ni a su emparejamiento con la distribución poblacional de trabajadores peruanos. Segundo, las correlaciones que aparecieron contraintuitivas no fueron exploradas por metodología local, esto es, con correlaciones que permiten identificar la fuerza de asociación en diferentes puntos de la variable (Bjerve, & Doksum, 1993), por ejemplo, en cada cuartil. Esto permitiría descomponer la covaribilidad en las regiones del constructo que pueden mantener diferentes intensidades de asociación y, por lo tanto, mejorar una estimación global de la asociación (Bjerve, & Doksum, 1993; Doksum, Blyth, Bradlow, Meng, & Zhao, 1994). Tercero, no se utilizaron medidas de asociación no lineal, y un examen más detallado de las relaciones bivariadas requerirá introducir este tipo de medidas para detectar con mejor exactitud el tamaño de estas asociaciones.

Figura 18

Factores Psicosociales del Trabajo y distrés (PHQ-4)

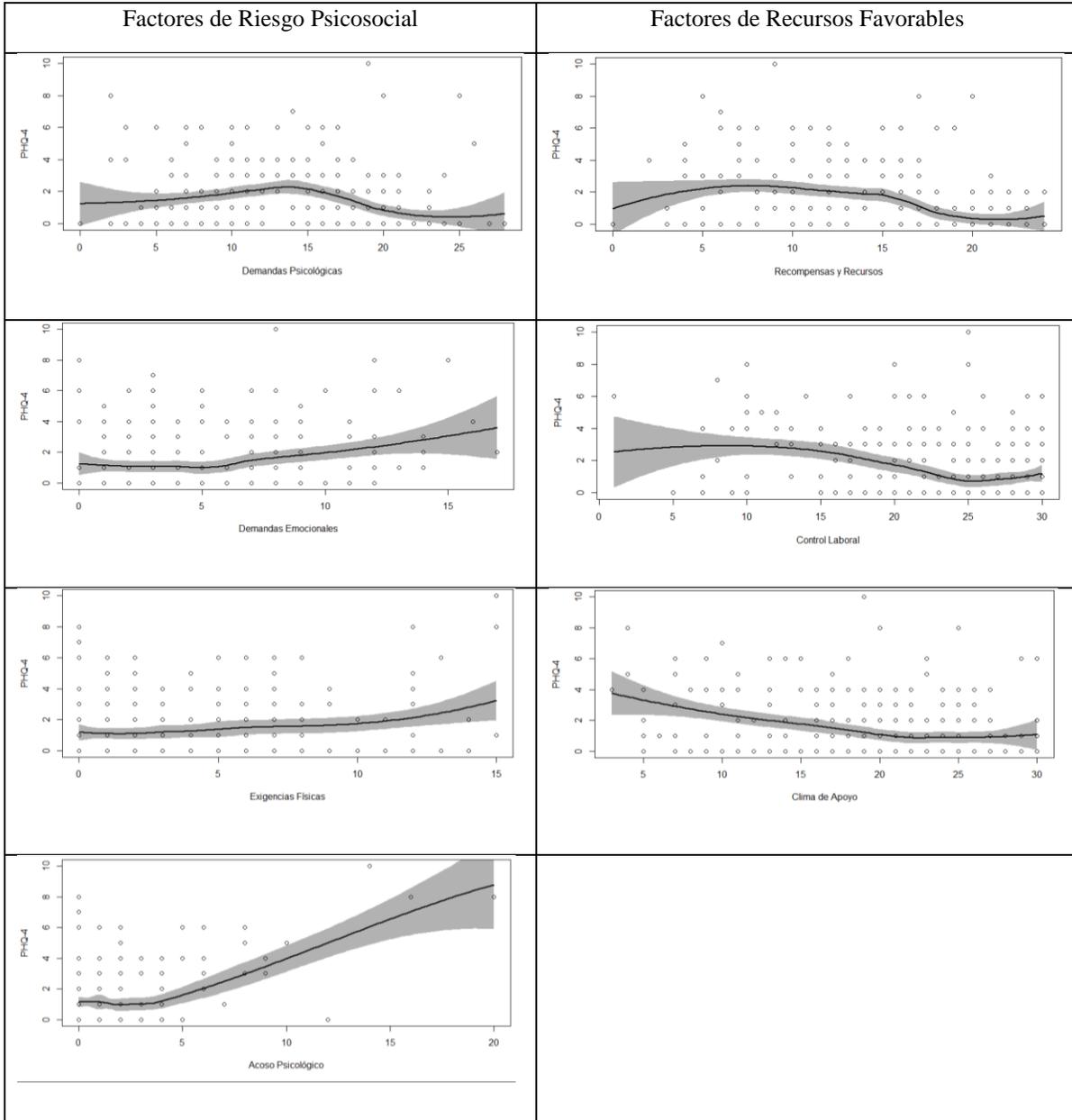


Figura 19

Factores Psicosociales del Trabajo y Estrés (SOS-S)

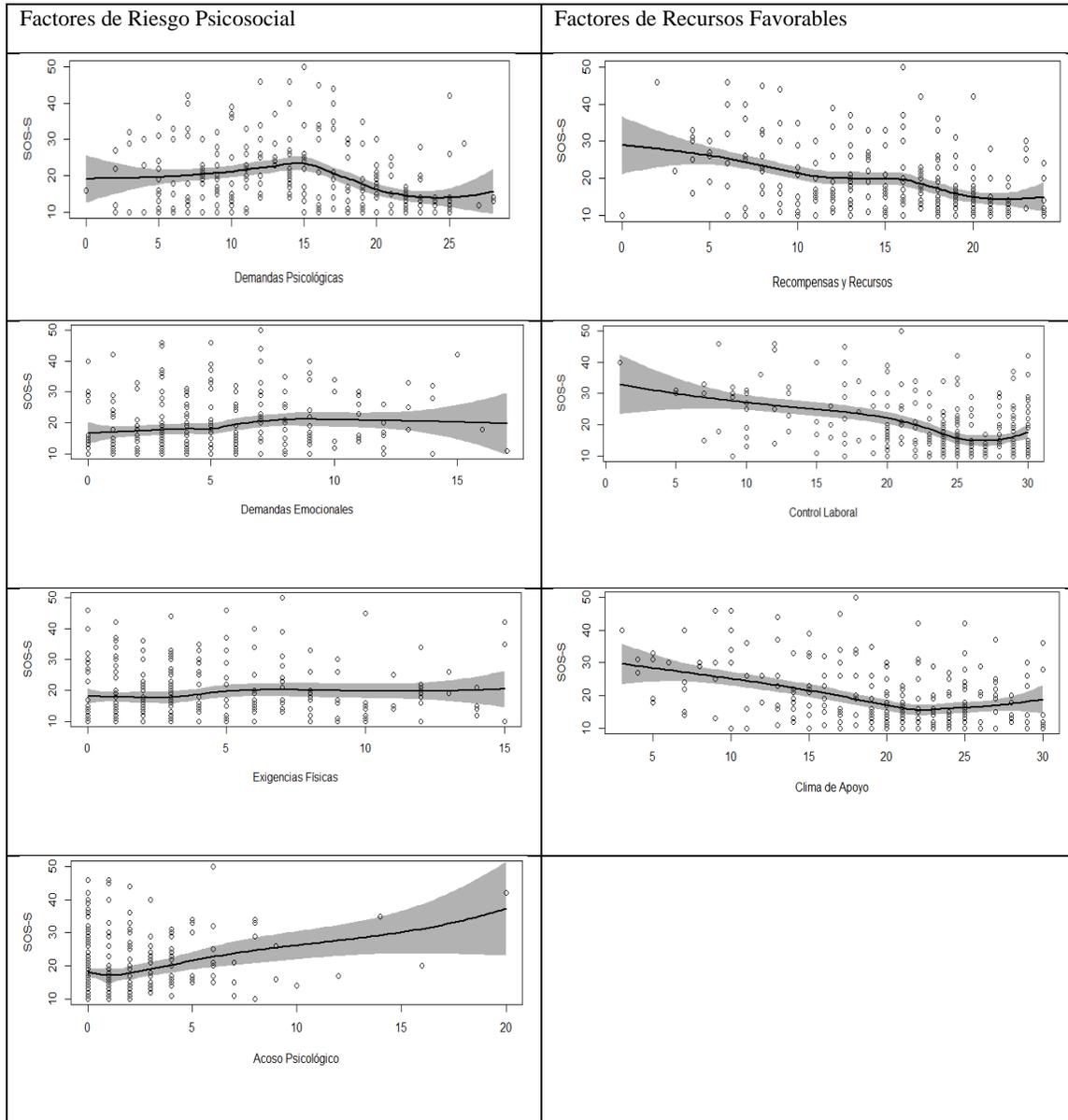


Figura 20

Factores Psicosociales del Trabajo y Estrés (ítem único)

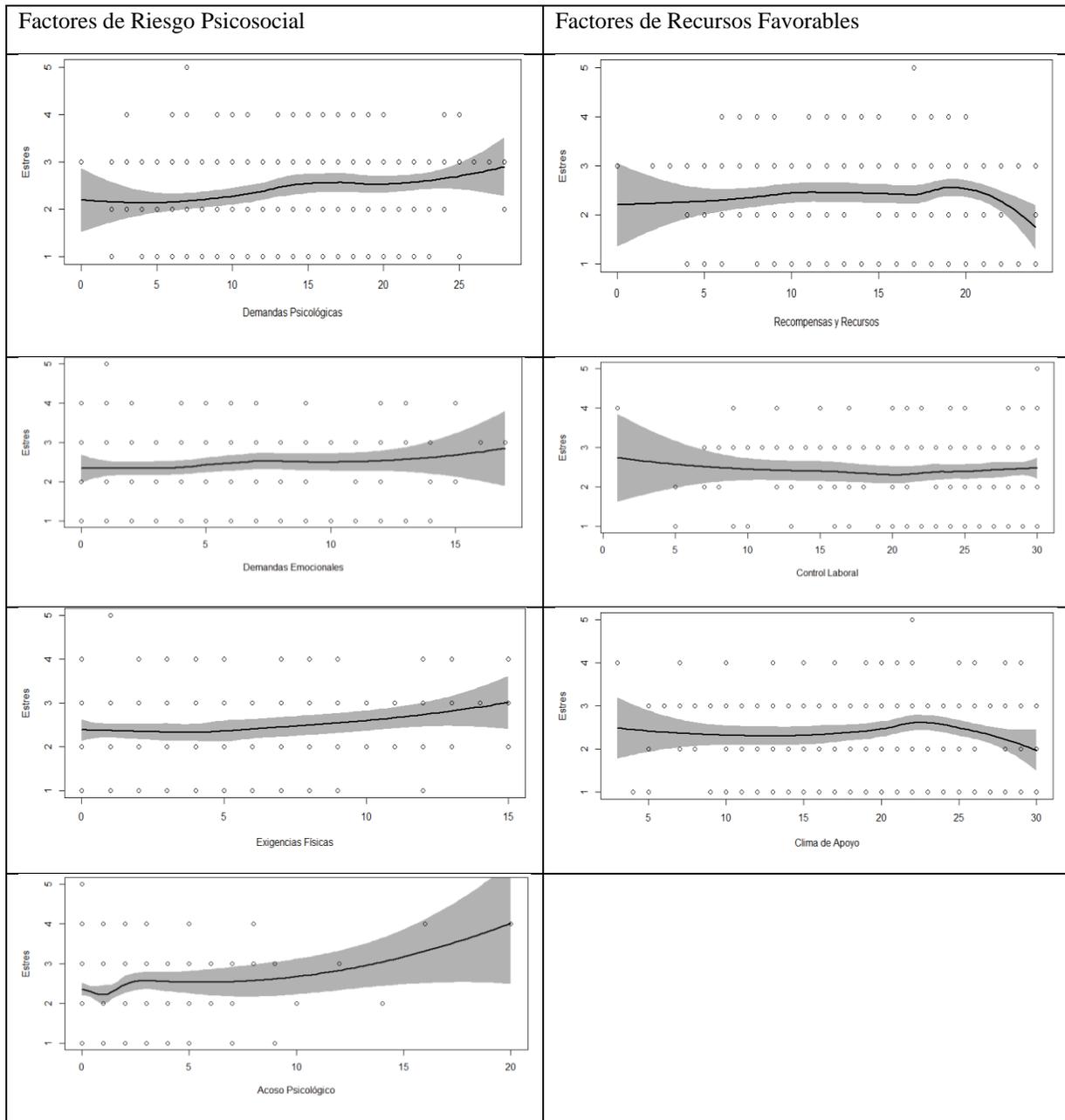


Figura 21

Factores Psicosociales del Trabajo y accidentes menores

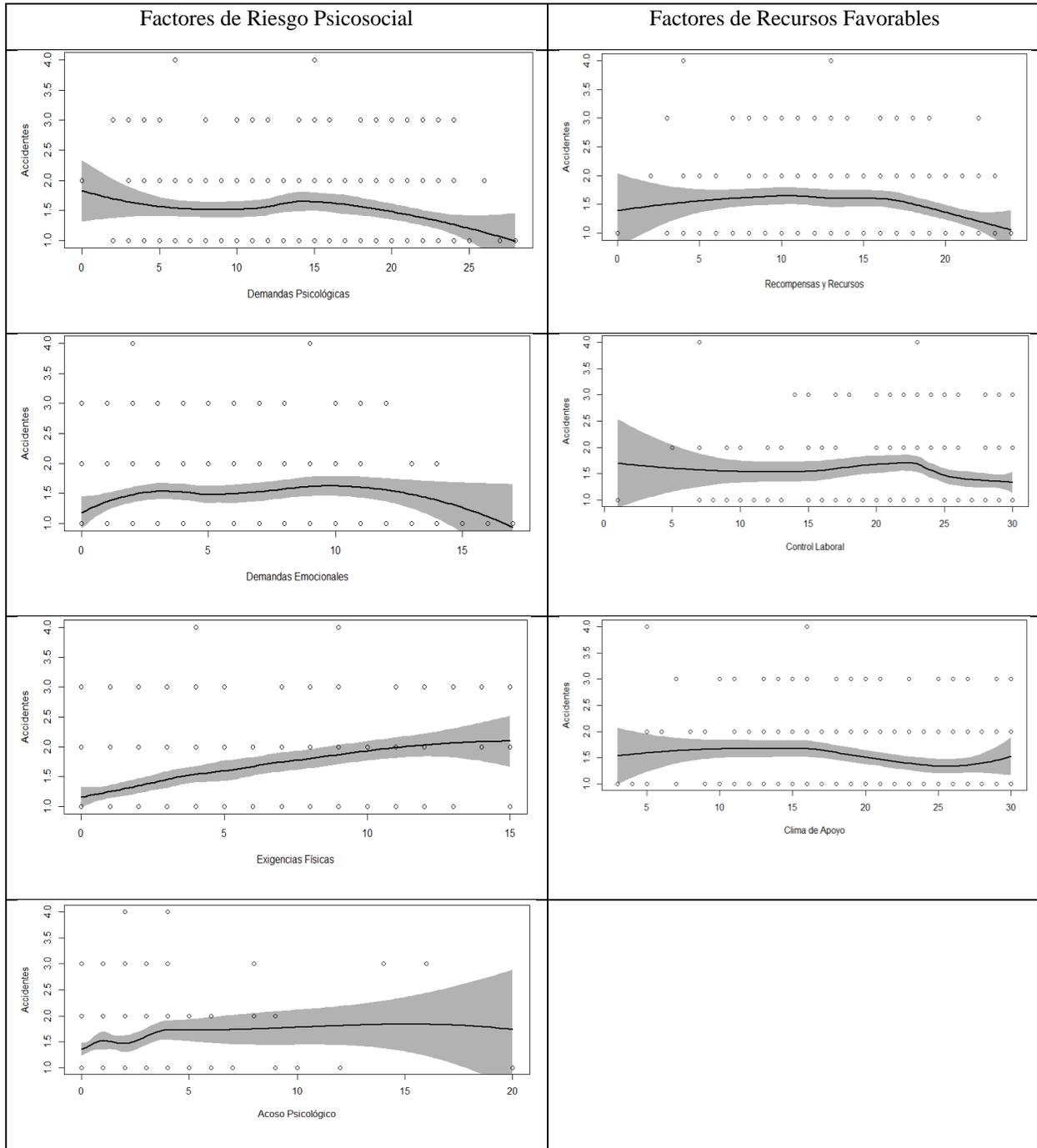


Figura 22

Factores Psicosociales del Trabajo y dificultades cognitivas

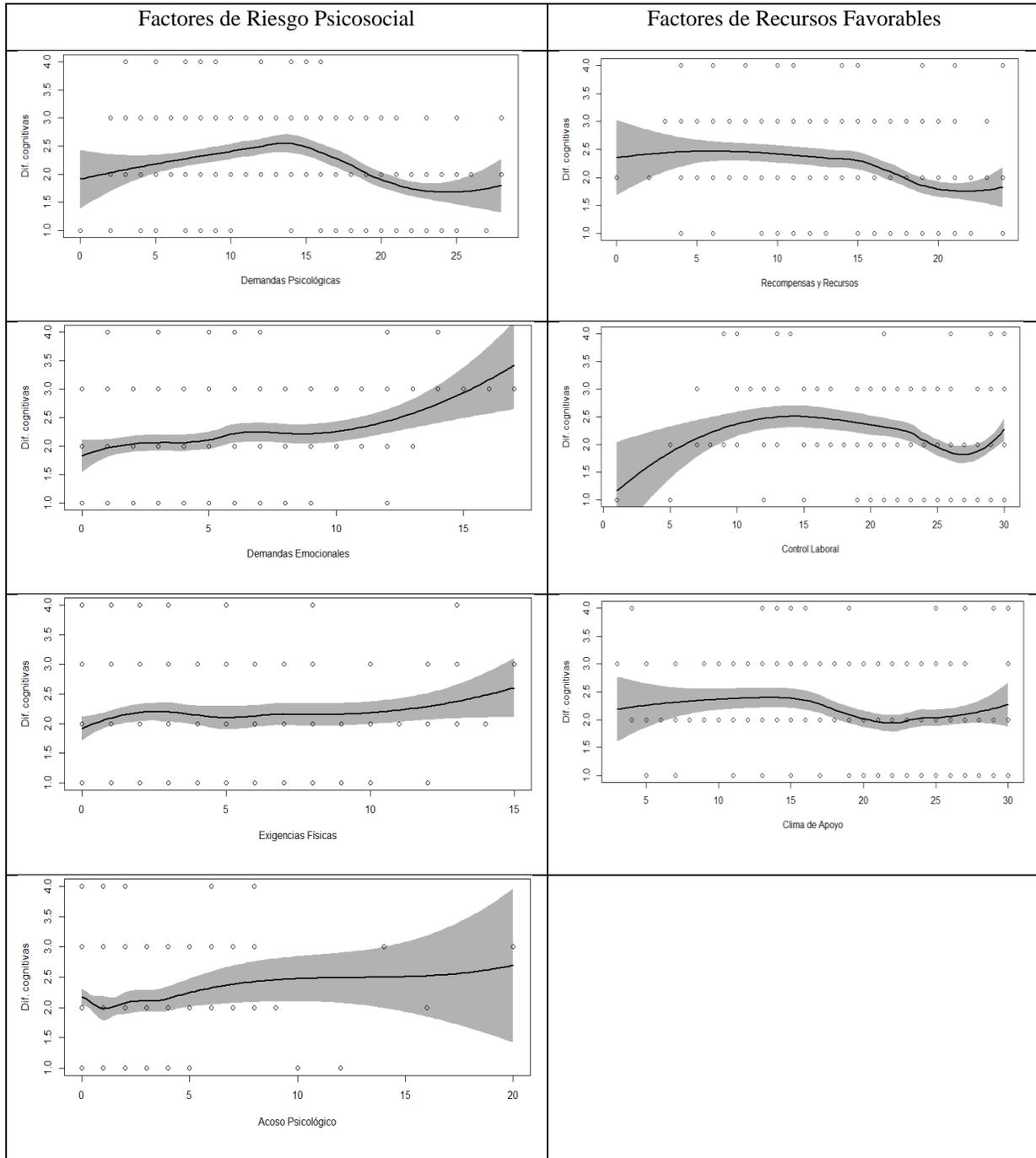
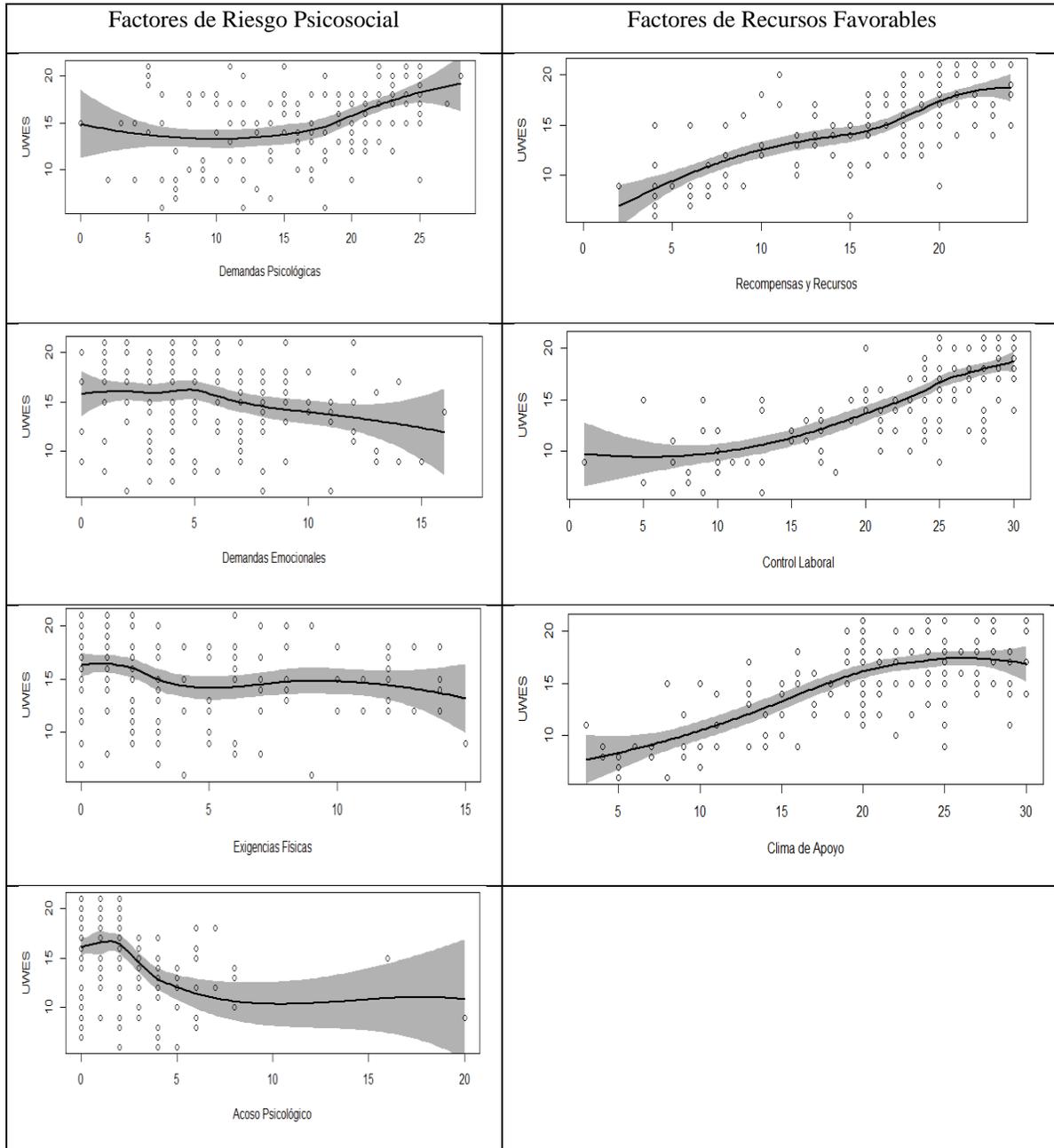


Figura 23

Factores Psicosociales del Trabajo y Engagement (UWES)



4. Puntos finales e implicaciones del proyecto

El proyecto realizado se orientó a adaptación una medida de factores psicosociales del trabajo, creada en México (PROPSIT) para trabajadores en ejercicio laboral contractual, hacia la población peruana de trabajadores que también ejercen laboralmente dentro de un marco contractual. El marco de adaptación fue la conceptualización general que los constructos evaluados por el PROPSIT tienen una fuerte característica emic (es decir, que son apropiados para la cultura de origen, pero también generalizables a otros contextos culturales), pero que, en un proceso de adecuación cultural, su contenido puede ser evaluado dentro de un proceso etic-emic, en que también se exploren contenidos emergentes que pueden ser propios en la cultura peruana. Para este propósito, el diseño del estudio fue nominado parcialmente mixto, una clasificación propuesta por Leech, & Onwuegbuzie (2009), que apropiadamente pondera el grado en que la metodología cualitativa y cuantitativa es combinada, es decir, que no se implementó un enfoque cualitativo-cuantitativo en todas las fases del estudio, sino en la primera fase (correspondiente a la validez de contenido) se implementó una nueva metodología mixta con componentes cualitativos y cuantitativos, mientras que, en el resto de fases, se implementó la metodología cuantitativa.

El marco para comprender y obtener las evidencias de validez fue el de los *Estandares*, un enfoque teórico para conceptualizar la validez de medida, especificando diferentes fuentes de evidencias de validez: de contenido, de procesos de respuesta, de estructura interna, y relación con otras variables. Esta visión del concepto de validez es influyente para la práctica de obtener evidencias de las propiedades psicométricas de una medida, y fue aparente que, en la literatura revisada, este marco no fue comúnmente utilizado para respaldar y dirigir investigaciones de validación de medidas de factores psicosociales. En la obtención de evidencias de validez de contenido, el presente estudio se diseñó para: a) identificar los estresores y motivadores laborales en la muestra peruana, bajo un marco a priori, b) estimar la claridad de los ítems, c) evaluar la relevancia conceptual de los ítems respecto a

sus dimensiones teóricas, y d) examinar la validez de los ítems con variables externas y de la estructura de respuesta del PROPSIT.

Respecto a la exploración de las experiencias con las situaciones estresantes y motivadores, se halló consistencia en el tipo categorial de ambas experiencias, porque en las tres preguntas inductoras, el ranking de las categorías identificadas en estas preguntas fue predominantemente muy alta. Esta consistencia interna no solo indica la reproducibilidad de la respuesta del trabajador, sino también su saliencia en la vivencia diaria durante y después de la jornada laboral. Una implicación es que, debido que la experiencia laboral es definicional del periodo humano de la adultez, la experiencia laboral definidas por sus estresores y motivadores debe ser un complemento clave de la salud mental del adulto. También se concluyó que las categorías conceptuales definidas por el modelo PROPSIT, fueron predominantemente apropiadas para identificar los factores psicosociales laborales de los trabajadores peruanos muestreados, lo cual añade fortaleza teórica y étic al modelo PROPSIT, creado en, y para, población laboral mexicana. Aunque algunas dimensiones conceptuales del PROPSIT (e.g., esfuerzo físico) no fueron reproducibles, esto parece estar asociados a las actividades predominantes de los trabajadores peruanos muestreados. En esta etapa del estudio, surgieron nuevas potenciales categorías que requieren más exploración (e.g., recreación como motivador laboral), con implicaciones para añadirlas en un modificación futura del cuestionario PROPSIT.

Respecto a la claridad de los ítems, los ítems del PROPSIT – FPS, individualmente o en conjunto, mostraron predominante lecturabilidad apropiada para trabajadores que como mínimo tuvieron secundaria básica. Luego de controlar el efecto de la potencial limitada competencia de lenguaje e identificar la falta de consenso entre los participantes, cada ítem fue percibido con alta claridad. El estudio tiene implicaciones metodológicas para evaluar la claridad de contenido, enfatizando la rigurosidad y la reducción de la práctica de usar solo un método para estimar los estadísticos de interés. Respecto a la relevancia conceptual por medio de trabajadores, todos los ítems

fueron percibidos como conceptualmente relevantes y definicionales en sus constructos propuestos. Este último aspecto evaluado, una implicación de que, con un procedimiento claro y eficiente, los participantes pueden ser una fuente de aproximación a los constructos medidos por un cuestionario. Esto puede ser vinculado con validez de *face* o validez sustantiva, en que se requiere la participación de los potenciales evaluados de un instrumento. Por otro lado, una implicación teórica, asociada a la corroboración de los constructos emic, es que los contenidos del PROPSIT fueron predominantemente reconocidos, y que corresponde a la importancia de estos constructos para configurar a los factores psicosociales desde un ángulo transcultural. Adicionalmente, los contenidos emergentes hallados sugieren nuevos constructos que requieren corroboración.

Respecto a la validez de los ítems con criterios externos del contenido del PROPSIT, se hallaron dos aspectos: primero, que las relaciones entre los ítems del PROPSIT - Recursos Favorables y las medidas externas mantienen asociaciones bajas o moderadas, pero todas con convergencia teórica; mientras que en los ítems del PROPSIT - factores de riesgo psicosocial hubo predominante convergencia teórica, pero existieron algunas discrepancias que pueden provenir de la idiosincrasia de la muestra, o de la baja intensidad de los factores de riesgo psicosocial muestreados por el PROPSIT. Y segundo, que la estructura de respuesta del PROPSIT puede requerir una modificación hacia un escalamiento de cinco opciones, para asegurar su eficiencia y adecuada representación del constructo. Una implicación de este estudio es que, la práctica estándar de adaptación de medidas debe incorporar la asociación con criterios de constructo o conductuales externos, más allá de la fuerza de la relación ítem-constructo usualmente obtenida con el modelamiento factorial. De este modo, la interpretación y la capacidad predictiva del constructo se fortalece a sus ítems.

En relación a las evidencias de procesos de respuesta, se identificaron los posibles sesgos en las respuestas al PROPSIT – FPST, enfocados en las respuestas descuidadas o con esfuerzo insuficiente. Se halló que este patrón afectó levemente a varios parámetros psicométricos como la replicabilidad del

constructo, las cargas factoriales, y la dimensionalidad. Este efecto, en general, fue positivo (i.e., incremento) de sus valores. Por otro lado, y sin un valor de referencia de prevalencia establecida, la prevalencia en la muestra estudiada puede considerarse moderada. Dos implicaciones pueden derivarse de este objetivo: primero, estos resultados sirvieron como referencia para los siguientes objetivos de este proyecto, porque los participantes que generaron estos patrones de respuesta fueron removidos. Segundo, la evaluación de la calidad de los datos debe ser parte del análisis del estudio de los factores psicosociales del trabajo, así como los métodos para su detección y tratamiento. Debido a su impacto sobre las estimaciones cuantitativas demostradas en el estudio y en la literatura en general, debe ser integrada en las buenas prácticas de análisis de datos en esta área.

En relación a la obtención de evidencias de estructura interna, los objetivos fueron la dimensionalidad, la invarianza de medición, la validez discriminativa, y la confiabilidad de los puntajes. El primero fue para identificar la estructura dimensional del PROPSIT, y fue explorada por procedimientos no paramétricos y paramétricos. Ambos métodos convergieron, y en que se requirieron dos iteraciones para definir la estructura dimensional en la muestra peruana. La primera iteración no mostró un ajuste insatisfactorio, mientras que la segunda tuvo un mejor ajuste, en que se removieron algunos ítems en ambas áreas, con base en análisis y en la teoría. El número y el contenido de ítems del modelo final en PROPSIT – factores de riesgo psicosocial y PROPSIT – factores de recursos favorables fue altamente similar con los ítems del estudio de dimensionalidad en muestra mexicana ($n = 1339$; Juárez-García & Flores-Jiménez, 2020). Tres implicaciones derivadas de este estudio son que: primero, la teoría relevante al PROPSIT adaptado en la muestra peruana adquiere parcialmente un valor emic, donde los constructos pueden ser vistos como generalizables; segundo, las dimensiones halladas fueron suficientemente fuertes y representativas de sus ítems, y no requirieron modificaciones adicionales que pueden representar el error de muestreo o una muestra idiosincrásica; tercero, la

incorporación de varios métodos cuyos resultados sean contrastados y evaluados en su convergencia o divergencia, ayuda a afinar las conclusiones obtenidas.

En el objetivo de la obtención de evidencias de validez convergente y divergente interna del PROPSIT, se halló que el PROPSIT en la muestra peruana cuenta con distintividad de los constructos con base teórica y empírica. Las correlaciones lineales entre las dimensiones no fueron grandes, de tal modo que no compromete la independencia estadística, y revela que los constructos pueden covariar en un grado que puede depender de la muestra y del contexto del uso del PROPSIT. Una implicación de esto es el fortalecimiento de la teoría subyacente del PROPSIT sobre la coexistencia de estos constructos, sin alta dependencia entre ellos, pero asociados. Por otro lado, la implicación metodológica es que el contraste de dos procedimientos habitualmente utilizados para este objetivo (criterio HTMT vs criterio Fornell-Larcker) ayudó a valorar las consecuencias de depender solo de una técnica.

En el objetivo sobre la equivalencia de medición o invarianza, se obtuvo que la estructura interna del PROPSIT-FPST (riesgos psicosociales) entre los grupos de trabajadores peruanos y mexicanos, balanceados en las características del sexo, edad y grado de instrucción, fue invariante respecto a la configuración de sus dimensiones (invarianza configuracional), la definición de cada dimensión (invarianza métrica) y la intensidad de respuesta de los ítems (invarianza de interceptos). Esta es una prueba importante y fundamental de la equivalencia de los constructos medidos en ambos grupos de trabajadores provenientes de dos contextos culturales diferentes. Tres implicaciones pueden reconocerse: una implicación de esto es que los constructos medidos pueden representar experiencias percibidas del ambiente laboral con similar conceptualización. Esta equivalencia ocurrió sobre una versión del PROPSIT-FPST que requirió remover algunos ítems. Segundo, que puede definirse una versión interculturalmente equivalente con este resultado, y que puede ser el estímulo para siguientes estudios de validación en otros contextos. Tercero, una implicación metodológica es que el

procedimiento de emparejamiento utilizado fue trascendental para aproximarse a una comparación razonable entre grupos de trabajadores con distintas ocupaciones.

En la evaluación de la confiabilidad de los puntajes del PROPSIT, se hallaron niveles que corresponden habitualmente con el uso de evaluaciones de grupo para descripción, y tomar decisiones apoyada en otras fuentes recolectadas. Por otro lado, el parecido entre los coeficientes α y ω , sugiere que ambos pueden ser aceptados como estimaciones de uso, especialmente el coeficiente alfa debido a su popularidad. En este estudio, sin embargo, la estimación fue más allá, y se obtuvo una aproximación a la confiabilidad condicional, donde se reveló que algunas regiones del puntaje pueden ser medidas con mejor precisión que en otros niveles del puntaje. Una implicación de este estudio es que el nivel de precisión de los puntajes puede requerir ir más allá de la estimación de coeficientes habituales o coeficientes recomendados, y de este modo tomar mejores decisiones sobre la elección y uso de los instrumentos. Otra implicación es que la estimación de dos coeficientes ayuda a comprender que esta propiedad no es dependiente solo de la “verdadera confiabilidad” de los puntajes, sino también del tipo de coeficiente utilizado, que para propósitos de investigación o de uso, puede ser recomendado aplicar ambos.

Finalmente, al examinar la relación de los constructos del PROPSIT – FPST con otras variables, en general, se halló consistencia teórica de los factores de riesgo psicosocial y factores de recursos favorables con constructos externos sobre reacciones emocionales (distrés psicológico y sobrecarga de estrés), indicadores subjetivos de rendimiento en el trabajo (accidentabilidad y dificultades cognitivas centradas en la memoria), y entusiasmo laboral. Sin embargo, una discrepancia importante fue hallada entre la dimensión de demandas psicológicas y el resto de variables externas, en donde el patrón de relaciones fue opuesto. Esto pudo ser un efecto de la intensidad de las demandas registradas en el grupo de participantes, es decir que los participantes no reportaron demandas frecuentemente muy intensas, y en consecuencia restringieron la mayor densidad de respuesta en las zonas moderadas o bajas en el

constructo. Estos niveles pueden ser suficientemente energizantes para dedicarse a las tareas laborales de manera eficiente, implicarse en el proceso del trabajo, y obtener gratificación por el desempeño en las tareas. Una implicación importante de este estudio es que las asociaciones entre constructos pueden ser detectadas no solo porque existen en los datos, sino también por la definición de esta relación; esta relación puede ser lineal o no lineal, y la elección del procedimiento de estimación de esta asociación debe ser coherente con la diferencia.

5. Referencias

- Abdelmoula, M., Chakroun, W., & Akrouf, F. (2015). The effect of sample size and the number of items on reliability coefficients: Alpha and ρ : a meta-analysis. *International Journal of Numerical Methods and Applications*, 13(1), 1-20.
http://doi.org/10.17654/IJNMAMar2015_001_020
- Abdin, E., Subramaniam, M., Chan, A., Chen, J. A., Chong, C. L., Wang, C., Lee, M., & Gan, S. L. (2019). iWorkHealth: An instrument to identify workplace psychosocial risk factors for a multi-ethnic Asian working population. *PLOS ONE*, 14(8), e0220566.
<https://doi.org/10.1371/journal.pone.0220566>
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955–959. <https://doi.org/10.1177/001316448004000419>
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45, 131-142. <https://doi.org/10.1177/0013164485451012>
- Allen, J., & Walsh, J. A. (2000). A construct-based approach to equivalence: Methodologies for cross-cultural/ multicultural personality assessment research. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment. Personality in clinical psychology series* (pp. 63-85). Mahway, NJ: Lawrence Erlbaum Associates.
- Almirall, P. & Hernández, J. (2017). Psicología y salud ocupacional I: Un recorrido histórico por el trabajo en el instituto nacional de salud de los trabajadores de la habana. *Revista Cubana de Salud y Trabajo*, 18(1), 57-66.
- Alsayed, A., & Manzi, G. (2019). A Comparison of Monotonic Correlation Measures with Outliers. *WSEAS Transactions on Computers*, 18, 223-230.

- Alves, M. G. M., Hökerberg, Y. H. M. & Faerstein, E. (2013). Trends and diversity in the empirical use of Karasek's demand-control model (job strain): a systematic review. *Revista Brasileira de Epidemiologia*, 16(1), 125-136. <https://doi.org/10.1590/s1415-790x2013000100012>
- American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME] (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Amirkhan, J.H. (2018). A brief stress diagnostic tool: The short stress overload scale. *Assessment*, 25(8), 1001–1013. <https://doi.org/10.1177/1073191116673173>
- Anderson, J. C. & Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.
- Andrea, H., Bültmann, U., van Amelsvoort, L. G. P. M., & Kant, Y. (2009). The incidence of anxiety and depression among employees—The role of psychosocial work characteristics. *Depression and Anxiety*, 26(11), 1040–1048. <https://doi.org/10.1002/da.20516>
- Angulo-Ramos, M., Córdova-Delgado, M., & Merino-Soto, C. (2021). Validez de contenido de un protocolo de Buenas Prácticas en la evaluación del desarrollo psicomotor. *Index De Enfermería*, 30(1-2), 100-104. <http://ciberindex.com/c/ie/e12872>
- Antoun, C., Conrad, F. G., Couper, M. P., & West, B. T. (2019). Simultaneous estimation of multiple sources of error in a smartphone-based survey. *Journal of Survey Statistics and Methodology*, 7, 93-117. <https://doi.org/10.1093/jssam/smy002>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *The American psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>

- Arapovic-Johansson, B., Wåhlin, C., Kwak, L., Björklund, C., & Jensen, I. (2017). Work-related stress assessed by a text message single-item stress question. *Occupational Medicine (Oxford, England)*, 67(8), 601–608. <https://doi.org/10.1093/occmed/kqx111>
- Arias, M. M., & Giraldo, C.V. (2011). El rigor científico en la investigación cualitativa. *Investigación y Educación en Enfermería*, 29, 501-514.
- Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, 52, 2489–2505. <https://doi.org/10.3758/s13428-020-01401-8>
- Asparouhov, T. & Muthen, B. O. (2006). Robust chi squared difference testing with mean and variance adjusted test statistics. Mplus Web Notes: No.10. Disponible en <http://www.statmodel.com/download/webnotes/webnote10.pdf>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397–438. <https://doi.org/10.1080/10705510903008204>
- Asparouhov, T., & Muthén, B. (2010). Simple second order chi-square correction. Disponible en http://www.statmodel.com/download/WLSMV_new_chi21.pdf
- Ato, M., López, J. J. & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Avallone F., & Paplomatas A. (2005). *Salute organizzativa: Psicologia del benessere nei contesti organizzativi*. Milano: Raffaello Cortina.
- Awai, N. S., Ganasegeran, K., & Abdul Manaf, M. R. (2021). Prevalence of workplace bullying and its associated factors among workers in a Malaysian public university hospital: A cross-sectional study. *Risk Management and Healthcare Policy*, 14, 75–85. <https://doi.org/10.2147/RMHP.S280954>

- Backé, E.-M., Seidler, A., Latza, U., Rossnagel, K. & Schumann, B. (2012). The role of psychosocial stress at work for the development of cardiovascular diseases: a systematic review. *International Archives of Occupational and Environmental Health*, 85, 67–79.
<https://doi.org/10.1007/s00420-011-0643-6>
- Bagozzi, R. P., Yi, Y., and Phillips, L. W. (1991). Assessing construct validity in organizational research. *Admin. Sci. Q.* 36, 421–458.
- Baker, R., Brick, M. J., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., & Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1(2), 90-143.
<https://doi.org/10.1093/jssam/smt008>
- Bakker, A. B. & Demerouti, E. (2007). The Job Demands-Resources Model: state of the art. *Journal of Managerial Psychology*, 22(3), 309-328. <http://doi.org/10.1108/02683940710733115>
- Bakker, A. B. & Rofríguez-Muñoz, A. (2012). Introducción a la psicología de la salud ocupacional positiva. *Psicothema*, 24(1), 62-65.
- Bakker, A. B., & Demerouti, E. (2007). The Job Demands-Resources model: State of the art. *Journal of Managerial Psychology*, 22(3), 309–328. <https://doi.org/10.1108/02683940710733115>
- Bakker, A. B., & Demerouti, E. (2014). Job Demands-Resources Theory. *Wellbeing*, 1–28.
<https://doi.org/10.1002/9781118539415.wbwell019>
- Bala, M., Chalil, G. R. B., & Gupta, A. (2012). Emic and *etic*: different lenses for research in culture: unique features of culture in Indian context. *Management and Labour Studies*, 37, 45 – 60.
<https://doi.org/10.1177/0258042X1103700105>
- Balducci, C., Avanzi, L., & Fraccaroli, F. (2014). Emotional demands as a risk factor for mental distress among nurses. *La Medicina del Lavoro*, 105(2), 100–108.

- Barling, J., & Griffith, A. (2003). A history of occupational health psychology. En: J. C. Quick y L.E. Tetrick, (Eds.), *Handbook of occupational health psychology* (pp. 19-33). Washington, DC: American Psychological Association.
- Barraza, A. (2011). Validación psicométrica de la escala unidimensional del burnout estudiantil. *Revista Intercontinental de Psicología y Educación, 13*(2), 51-74.
- Barrio, I. (2008). Validación de la Escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes. *Anales del Sistema Sanitario de Navarra, 31*(2), 135-152.
- Battaglia, M. (2008a). Nonprobability sampling. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 523-526). Thousand Oaks, CA: SAGE Publications, Inc.
<https://doi.org/10.4135/9781412963947.n337>
- Battaglia, M. (2008b). Purposive sample. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 645-647). Thousand Oaks, CA: SAGE Publications, Inc.
<https://doi.org/10.4135/9781412963947.n419>
- Beaton, D. E., Bombardier, C., Guillemin, F., Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186-3191.
- Bentahar, O., & Cameron, R. (2015). Design and Implementation of a Mixed Method Research Study in Project Management. *The Electronic Journal of Business Research Methods, 13*(1), 3-15.
- Bentler, P. M., & Dugeon, P. (1996). Covariance structure Analysis: statistical practice, theory, and directions. *Annual Review of Psychology, 47*, 563-592.
<https://doi.org/10.1146/annurev.psych.47.1.563>
- Bentler, P. M., & Wu, E. J. C. (2012). *EQS for Windows*. Encino CA: Multivariate Software, Inc.
- Bentler, P. M. (1986). *Lagrange Multiplier and Wald Tests for EQS and EQS/PC*. Los Angeles: BMDP Statistical Software.

- Bergkvist, L. (2014). Appropriate use of single-item measures is here to stay. *Marketing Letters*, 26(3), 245–255. <https://doi.org/10.1007/s11002-014-9325-y>
- Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple item versus single-item technology of the same constructs. *Journal of Marketing Research*, 44, 175-184. <https://doi.org/10.1509/jmkr.44.2.175>
- Berry, J. (1969). On cross-cultural comparability. *International Journal of Psychology*, 4, 119 – 128. <https://doi.org/10.1080/00207596908247261>
- Berry, J. (1989). Imposed etics – emics – derived etics: The operationalisation of a compelling idea. *International Journal of Psychology*, 24, 721 – 735. <https://doi.org/10.1080/00207598908246808>
- Berry, J.W. (1990). Imposed etics, emics, derived etics: Their conceptual and operational status in cross-cultural psychology. In Headland, T.N., Pike, K.L., Harris, M. (Eds), *Emics and etics: The insider/outsider debate* (pp. 28–47). Newbury Park, CA: Sage.
- Berthelsen, H., Westerlund, H., Bergström, G., & Burr, H. (2020). Validation of the Copenhagen Psychosocial Questionnaire Version III and Establishment of Benchmarks for Psychosocial Risk Management in Sweden. *International Journal of Environmental Research and Public Health*, 17(9), 3179. <http://doi.org/10.3390/ijerph17093179>
- Berthelsen, M., Pallesen, S., Magerøy, N., Tyssen, R., Bjorvatn, B., Moen, B. E., & Knardahl, S. (2015). Effects of psychological and social factors in shiftwork on symptoms of anxiety and depression in nurses: A 1-year follow-up. *Journal of Occupational and Environmental Medicine*, 57(10), 1127–1137. <https://doi.org/10.1097/JOM.0000000000000532>
- Besser, A., Priel, B., Flett, G. L., & Wiznitzer, A. (2007). Linear and nonlinear models of vulnerability to depression: Personality and postpartum depression in a high risk population. *Individual Differences Research*, 5(1), 1–29.

- Bjerve, S., & Doksum, K. (1993). Correlation curves: Measures of association as functions of covariate values. *The Annals of Statistics*, 21(2), 890–902. <http://www.jstor.org/stable/2242266>
- Black, K. (2010). *Business Statistics: Contemporary Decision Making* (6th edition). Hoboken, NJ: John Wiley & Sons
- Bletzer K. V. (2015). Visualizing the qualitative: making sense of written comments from an evaluative satisfaction survey. *Journal of Educational Evaluation for Health Professions*, 12, 12. <https://doi.org/10.3352/jeehp.2015.12.12>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Boehnke, K., Lietz, P., Schreier, M., & Wilhelm, A. (2011). Sampling: The selection of cases for culturally comparative psychological research. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 101-129). New York, NY: Cambridge University Press.
- Bono, R., Blanca, M. J., Arnau, J., & Gómez-Benito, J. (2017). Non-normal Distributions Commonly Used in Health, Education, and Social Sciences: A Systematic Review. *Frontiers in psychology*, 8, 1602. <https://doi.org/10.3389/fpsyg.2017.01602>
- Boone W. J., & Staver J. R. (2020). Presentation and explanation techniques to use in Rasch articles. In: *Advances in Rasch Analyses in the Human Sciences*. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-43420-5_19
- Boschi, H., Trenoweth, S., & Sheppard, Z. A. (2017). Stress at work: Factors associated with cognitive disorganisation among private sector professionals. *Health Psychology Open*, 1-8. <https://doi.org/10.1177/2055102917718376>

- Boulianne, N., & Fortin, E. (2006). Conditions de travail et santé: les enquêtes statistiques au Québec. *Santé, Société et Solidarité*, 2, 143-146. <https://doi.org/10.3406/oss.2006.1134>
- Bowling, A. (2005). Just one question: If one question works, why ask several? *Journal of Epidemiology and Community Health*, 59(5), 342–345. <https://doi.org/10.1136/jech.2004.021204>
- Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.
- Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, 22(3), 450–466. <https://doi.org/10.1037/met0000074>
- Broetje, S., Jenny, G. J., & Bauer, G. F. (2020). The key job demands and resources of nursing staff: an integrative review of reviews. *Frontiers in psychology*, 11, 84. <https://doi.org/10.3389/fpsyg.2020.00084>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.
- Browne, M. (1972). Oblique rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 207-212.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111-150. https://doi.org/10.1207/S15327906MBR3601_05
- Buil, I., de Chernatony, L. & Martínez, E. (2012). Methodological issues in cross-cultural research: An overview and recommendations. *Journal of Targeting, Measurement and Analysis for Marketing*, 20(3-4), 223-234. <https://doi.org/10.1057/jt.2012.18>
- Bültmann, U., de Vries, M., Beurskens, A. J., Bleijenberg, G., Vercoulen, J. H., & Kant, I. (2000). Measurement of prolonged fatigue in the working population: determination of a cutoff point for the checklist individual strength. *Journal of occupational health psychology*, 5(4), 411–416. <https://doi.org/10.1037//1076-8998.5.4.411>

- Burchett, D., Dragon, W. R., Smith Holbert, A. M., Tarescavage, A. M., Mattson, C. A., Handel, R. W., & Ben-Porath, Y. S. (2016). “False feigners”: Examining the impact of non-content-based invalid responding on the Minnesota Multiphasic Personality Inventory-2 Restructured Form content-based invalid responding indicators. *Psychological Assessment, 28*(5), 458–470. <https://doi.org/10.1037/pas0000205>
- Burke, M. J., & Dunlap, W. P. (2002). Estimating interrater agreement with the average deviation index: a user’s guide. *Organizational Research Methods, 5*(2), 159–172. <https://doi.org/10.1177/1094428102005002002>
- Burman, R., & Goswami, T. G. (2018). A Systematic Literature Review of Work Stress. *International Journal of Management Studies, 3*(9), 112-132. [http://dx.doi.org/10.18843/ijms/v5i3\(9\)/15](http://dx.doi.org/10.18843/ijms/v5i3(9)/15)
- Byrne, B. (2016). Adaptation of assessment scales in cross-national research: issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation, 5*(1), 51–65. <http://dx.doi.org/10.1037/ipp0000042>
- Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*(4), 872-882.
- Calderón-de la Cruz, G., Domínguez, Sergio., & Arroyo-Rodríguez, Fiorella. (2018). Análisis psicométrico preliminar de una medida breve de autoeficacia profesional en trabajadores peruanos: AU-10. *Psicogente, 21*(39), 12-24. <https://doi.org/10.17081/psico.21.39.2819>
- Calderón-De la Cruz, G., Merino-Soto, C., Juarez-Garcia, A. Jimenez-Clavijo, M. (2018). Propiedades psicométricas de la escala de Carga de trabajo en trabajadores peruanos. *Archivos de Prevención de Riesgos Laborales, 21*(3), 123-127. <http://dx.doi.org/10.12961/apr1.2018.21.03.2>
- Calderon-De la Cruz, G., Merino-Soto, C., Reyes-Ayala, I., & Luna-Ayón, K. (2021, en prensa). Versión ultra reducida del UWES-3Utrech Work Engagement: Validación preliminar en trabajadores peruanos. *Archivos de Prevención de Riesgos Laborales*.

- Camacho, A. (2019). Estrés y salud en trabajadores obreros: Un estudio mixto. Tesis para optar el grado de Doctor, Universidad Autónoma del Estado de Morelos, Cuernavaca, México.
- Campos, J. A. D. B., Zucoloto, M. L., Bonafé, F. S. S., Jordani, P. C., & Maroco, J. (2011). Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration. *Computers in Human Behavior*, 27(5), 1875-1883. <https://doi.org/10.1016/j.chb.2011.04.011>
- Charria, V. H., Sarsosa, K. & Arenas, F. (2011). Factores de riesgo psicosocial laboral: métodos e instrumentos de evaluación. *Revista Facultad Nacional de Salud Pública*, 29(4), 380-391.
- Charter, R. A. (1999.) Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21(4), 559-566. <https://doi.org/10.1076/jcen.21.4.559.889>
- Charter, R. A. (2001). Damn the precision, full speed ahead with the clinical interpretation. *Journal of Clinical and Experimental Neuropsychology*, 23(5), 692-694. <https://doi.org/10.1076/jcen.23.5.692.1237>
- Chaudhary, R. (2014). Occupational self efficacy expectations among Indian executives: Examining the psychometric properties of Occupational Self Efficacy Scale (OSES). *Global Business Review*, 15(1), 47–58. <https://doi.org/10.1177/0972150913515603>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Chen, P.Y. & Popovich, P.M. (2002). *Correlation: Parametric and nonparametric measures*. Thousand Oaks. CA: Sage. <https://doi.org/10.4135/9781412983808>
- Cheung, G. W. & Wang, C. (2017). Current approaches for assessing convergent and discriminant validity with sem: issues and solutions. *Academy of Management Annual Meeting Proceedings*, 2017. <https://doi.org/10.5465/AMBPP.2017.12706abstract>

- Cheung, G.W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233 – 255.
https://doi.org/10.1207/s15328007sem0902_5
- Chungkham, H. S., Ingre, M., Karasek, R., Westerlund, H., & Theorell, T. (2013). Factor structure and longitudinal measurement invariance of the demand control support model: an evidence from the Swedish Longitudinal Occupational Survey of Health (SLOSH). *PloS one*, 8(8), e70541.
<https://doi.org/10.1371/journal.pone.0070541>
- Clark, M. (2013). A comparison of correlation measures. Center for Social Research, University of Notre Dame, 4. Disponible en:
<http://www3.nd.edu/~mclark19/learn/CorrelationComparison.pdf>
- Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. In J. M. Chambers, & T. J. Hastie (Eds.), *Statistical Models in S* (p.p. 309–376). New York: Chapman & Hall.
<https://doi.org/10.1201/9780203738535>
- Cohen, A. (1991). Career stage as a moderator of the relationships between organizational commitment and its outcomes: A meta-analysis. *Journal of Occupational Psychology*, 64, 253-268.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cohen, J. (2008). *Statistical power analysis for the behavioral sciences*. Second Edition. Hillsdale, New Jersey: Lawrence Erlbaum Associatesm Inc.
- Conway P. M. (2009). Gli strumenti per la valutazione soggettiva del rischio stress lavoro-correlato [The instruments for subjective assessment of work-related stress]. *Giornale Italiano di Medicina del Lavoro ed Ergonomia*, 31(2), 197–199.
- Conway, P. (2009). The instruments for subjective assessment of work-related stress. *Giornale Italiano di Medicina del Lavoro ed Ergonomia*, 31, 197–199.

- Cooper, H. (2020). *Reporting quantitative research in psychology: How to meet APA style journal article reporting standards (2nd Ed.)*. Washington, DC: American Psychological Association.
- Coppock, A. (2018). Generalizing from survey experiments conducted on mechanical Turk: A replication approach. *Political Science Research and Methods*, 7(3), 613-628.
<http://doi.org/10.1017/psrm.2018.10>
- Coppock, A., Leeper, T. J., & Mullinix, K. J. (2018). Generalizability of heterogeneous treatment effect estimates across samples. *Proceedings of the National Academy of Sciences*, 115(49), 12441-12446. <http://doi.org/10.1073/pnas.1808083115>
- Correll, J., Mellinger, C., McClelland, G. H., & Judd, C. M. (2020). Avoid Cohen's 'small', 'medium', and 'large' for power analysis. *Trends in cognitive sciences*, 24(3), 200–207.
<https://doi.org/10.1016/j.tics.2019.12.009>
- Costello, A. B. & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7). <http://pareonline.net/getvn.asp?v=10&n=7>
- Coulacoglou, C., & Saklofske, D. H. (2017). *Psychometrics and Psychological assessment*. San Diego CA: Academic Press.
- Creswell, J. W. (2013). *Research designs: Qualitative, quantitative, and mixed methods approaches (4th ed.)*. Thousand Oaks, CA: Sage.
- Crişan, D. R., Tendeiro, J. N., & Meijer, R. R. (2020). On the practical consequences of misfit in Mokken scaling. *Applied Psychological Measurement*, 44(6), 482–496.
<https://doi.org/10.1177/0146621620920925>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>

- Cronk, B C., & West, J. L. (2002). Personality research on the Internet: A comparison of Web-based and traditional instruments in take-home and in-class settings. *Behavior Research Methods, Instruments, & Computers*, 34(2), 177-180. <https://doi.org/10.3758/bf03195440>
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4) 349-354.
<https://doi.org/10.1037/h0047358>
- Crutzen, R., & Peters, G.-J. Y. (2017). Scale quality: Alpha is an inadequate estimate and factor-analytic evidence is needed first of all. *Health Psychology Review*, 11, 242–247.
<https://psycnet.apa.org/doi/10.1080/17437199.2015.1124240>
- de Siqueira Santos, S., Takahashi, D. Y., Nakata, A., & Fujita, A. (2014). A comparative study of statistical methods used to identify dependencies between gene expression signals. *Briefings in Bioinformatics*, 15(6), 906–918. <https://doi.org/10.1093/bib/bbt051>
- Demerouti, E., & Bakker, A. B. (2011). The Job Demands– Resources model: Challenges for future research. *SA Journal of Industrial Psychology*, 37(2), 1-9.
<http://doi.org/10.4102/sajip.v37i2.974>
- Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, 86, 499–512. <https://doi.org/10.1037//0021-9010.86.3.499>
- Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and psychological measurement*, 77(2), 185–203.
<https://doi.org/10.1177/0013164416658325>
- Denscombe, M. (2009). Item non-response rates: a comparison of online and paper questionnaires. *International Journal of Social Research Methodology*, 12(4), 281-291.
<https://doi.org/10.1080/13645570802054706>

- DeSalvo, K. B., Fisher, W. P., Tran, K., Bloser, N., Merrill, W., & Peabody, J. (2006). Assessing measurement properties of two single-item general health measures. *Quality of Life Research, 15*(2), 191-201. <https://doi.org/10.1007/s11136-005-0887-2>
- Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science, 40*(3), 434–449. <https://doi.org/10.1007/s11747-011-0300-3>
- DiCiccio, T. J. & Efron, B. (1996). Bootstrap confidence intervals (with Discussion). *Statistical Science, 11*, 189–228.
- Dima, A. L. (2018) Scale validation in applied health research: tutorial for a 6-step R-based psychometrics protocol. *Health Psychology and Behavioral Medicine, 6*(1), 136-161. <https://doi.org/10.1080/21642850.2018.1472602>
- Dixon, P.M. (2006). Bootstrap resampling. In: A.H. El-Shaarawi, W.W. Piegorsch and G. Høst (Eds.), *Encyclopedia of Environmetrics* (pp. 212-220). New York: Wiley. <https://doi.org/10.1002/9780470057339.vab028>
- Doksum, K., Blyth, S., Bradlow, E., Meng, X., & Zhao, H. (1994). Correlation curves as local measures of variance explained by regression. *Journal of the American Statistical Association 89*(426), 571–582. <http://www.jstor.org/stable/2290860>
- Dorociak, K. E., Rupert, P. A., & Zahniser, E. (2017). Work life, well-being, and self-care across the professional lifespan of psychologists. *Professional Psychology: Research and Practice, 48*(6), 429-437. <http://dx.doi.org/10.1037/pro0000160>
- Duan, W., & Mu, W. (2018). Validation of a Chinese version of the stress overload scale-short and its use as a screening tool for mental health status. *Quality of life research: an International*

Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, 27(2), 411–421.

<https://doi.org/10.1007/s11136-017-1721-3>

Dümcke, S., Mansmann, U., & Tresch, A. (2014). A novel test for independence derived from an exact distribution of *i*th nearest neighbours. *PLOS ONE*, 9(10): e107955.

<https://doi.org/10.1371/journal.pone.0107955>

Dunn, A.M., Heggstad, E.D., Shannock, L.R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, 33(1), 105-121.

<https://doi.org/10.1007/s10869-016-9479-0>

Duru, P., Ocaktan, M. E., Çelen, Ü., & Örsal, Ö. (2018). The effect of workplace bullying perception on psychological symptoms: a structural equation approach. *Safety and Health at Work*, 9(2),

210-215. <https://doi.org/10.1016/j.shaw.2017.06.010>

Edwards, J. E., Thomas, M. D., Rosenfeld, P., & Booth-Kewley, S. (1997). *How to conduct organizational surveys: A step-by-step guide*. Thousand Oaks, CA: Sage.

Edwards, J., & Webster, S. (2012). Psychosocial risk assessment: Measurement invariance of the UK Health and Safety Executive's Management Standards Indicator Tool across public and private sector organizations. *Work & Stress*, 26, 130–142.

<https://doi.org/10.1080/02678373.2012.688554>

Elliott, M. R., & Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, 32(2), 249-264.

Elo, A.-L., Leppänen, A. & Jahkola, A. (2003). Validity of a single-item measure of stress symptoms. *Scandinavian Journal of Work Environmental Health*, 29(6), 444-451.

<https://doi.org/10.5271/sjweh.752>

- Engelhard, G., & Wind, S. A. (2012, April). Rating quality studies using Rasch measurement theory. Paper presented at the annual meeting of the American Educational Research Association in Vancouver (April 2012).
- Eskildsen, A., Andersen, L. P., Pedersen, A. D., Vandborg, S. K., & Andersen, J. H. (2015). Work-related stress is associated with impaired neuropsychological test performance: A clinical cross-sectional study. *Stress, 18*(2), 198–207. <https://doi.org/10.3109/10253890.2015.1004629>
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1-4. <http://doi.org/10.11648/j.ajtas.20160501.11>
- Farrell, A. M., & Rudd, J. M. (2009). Factor analysis and discriminant validity: a brief review of some practical issues. In: D. Tojib (Ed.), *ANZMAC 2009 conference proceedings*. Australia: ANZMAC. Disponible en: https://publications.aston.ac.uk/id/eprint/7644/1/factor_analysis_ANZMAC_2009.pdf
- Feldt, L. S., Steffen, M., & Gupta, N. C. (1985). A comparison of five methods for estimating the standard error of measurement at specific score levels. *Applied Psychological Measurement, 9*(4), 351–361. <https://doi.org/10.1177/014662168500900402>
- Fellows, I. (2018). wordcloud: Word Clouds. R package version 2.6. <https://CRAN.R-project.org/package=wordcloud>
- Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods, 80–92*. <https://doi.org/10.1177/160940690600500107>
- Fernandes, C., & Pereira, A. (2016). Exposure to psychosocial risk factors in the context of work: a systematic review. *Revista de Saúde Pública, 50*, 24. <http://doi.org/10.1590/S1518-8787.2016050006129>

- Fernández Huerta, J. (1959). Medidas sencillas de lecturabilidad. *Consigna*, 214, 29-32.
- Fernández, M. & Merino, C. (2014). Error de medición alrededor de los puntos de corte en el MBI-GS. *Liberabit*, 20(2), 209-218.
- Fernández, M., Juárez, A. & Merino, C. (2015). Invarianza de medición y estructural del MBI-GS en una muestra multiocupacional peruana. *Liberabit*, 21, 9-20.
- Fernández-Arata, M. & Calderón-de la Cruz, G. & Navarro-Loli, J. (2016). Psicología de la Salud Ocupacional: Una especialidad emergente en el Perú. *Revista Herediana de Medicina*, 28, 281-282
- Fernández-Arata, M. & Calderón-de la Cruz, G. (2017). Modelo Demandas-Control-Apoyo social en el estudio del estrés laboral en el Perú. *Revista Herediana de Medicina*, 28, 281-282.
- Fernández-Arata, M. & Merino-Soto, C. (2016). El lado socialmente deseable de las respuestas a medidas de burnout y engagement: un estudio preliminar. *Revista Colombiana de Psicología*, 25(1), 83-94. <https://doi.org/10.15446/rcp.v25n1.47648>
- Fernández-López, J. A., Fernández-Hidalgo, E. & Siegrist, J. (2005). El trabajo y sus repercusiones en la salud: el modelo “Desequilibrio Esfuerzo-Recompensa-DER”. *Revista de Calidad Asistencial*, 20(3), 165-170. [https://doi.org/10.1016/s1134-282x\(08\)74743-2](https://doi.org/10.1016/s1134-282x(08)74743-2)
- Ferrando, P. J. & Lorenzo-Seva, U. (2014). El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales. *Anales de Psicología*, 30(3), 1170-1175
<https://doi.org/10.6018/analesps.30.3.199991>
- Ferrando, P. J. (1999). Likert scaling using continuous, censored, and graded response models: effects on criterion-related validity. *Applied Psychological Measurement*, 23, 161–175.
<https://doi.org/10.1177/01466219922031284>

- Ferrando, P. J., & Lorenzo-Seva, U. (2018). Assessing the quality and appropriateness of factor solutions and factor score estimates in exploratory item factor analysis. *Educational and Psychological Measurement*, 78, 762–780. <https://doi.org/10.1177/0013164417719308>
- Ferrario, M., Fornari, C., Borchini, R., Merluzzi, F., & Cesana, G. (2005). La percezione dello stress nel terziario: utilità e indicazioni del Job Content Questionnaire [Job stress in the service industry. Evaluation of the Italian version of the Job Content Questionnaire]. *La Medicina del Lavoro*, 96(3), 191–199.
- Fetvadjev, V. H., & Van de Vijver, F. J. R. (2017). Cross-cultural research. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–12). New York, NY: Springer. https://doi.org/10.1007/978-3-319-28099-8_1294-1
- Figueiredo-Ferraz, H.; Gil-Monte, P. R.; Olivares-Faundez, V. E. (2015). Influence of mobbing (workplace bullying) on depressive symptoms: a longitudinal study among employees working with people with intellectual disabilities. *Journal of Intellectual Disability Research*, 59(1), 39–47. <https://doi.org/10.1111/jir.12084>
- Finch, W. H. (2012). Distribution of variables by method of outlier detection. *Frontiers in Psychology*, 3, 211. <https://doi.org/10.3389/fpsyg.2012.00211>
- Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In Hancock, G. R., Mueller, R. O. (Eds.), *A Second Course in Structural Equation Modeling* (2nd ed., pp. 439–492). Charlotte, NC: Information Age.
- Finstad, K. (2010). Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales. *Journal of Usabilidad Studies*, 5(3), 104–110.
- Fisher, R., Fontaine, J., van de Vijver, F., & van Hemert, D. A. (2009). An examination of acquiescent response styles in cross-cultural research. In A. Gari, & K. Mylonas (Ed.s) *Quod erat*

Demonstrandum: From Herodotus' Ethnographic Journeys to Cross-Cultural Research (pp. 137-148). Athens: Pedio Books Publishing.

Fleming, J. S., & Merino Soto, C. (2005). Medidas de simplicidad y de ajuste factorial: un enfoque para la evaluación de escalas construidas factorialmente. *Revista De Psicología*, 23(2), 250-266.
<https://doi.org/10.18800/psico.200502.002>

Flores, C., & Juárez, A. (2016). Factores psicosociales y síndrome de Burnout en instructores comunitarios: una aproximación desde un análisis mixto. *Revista Mexicana de Salud en el Trabajo*, 7-8(17-18), 3-9.

Flores, C., Fernández, M., Merino, C., Guimet, M., & Juárez, A. (2015). Entusiasmo por el trabajo (Engagement): Un estudio de validez en Profesionales de la docencia en Lima, Perú. *Liberabit*, 21(2), 195-206.

Flores-Jimenez, C. A., Merino-Soto, C., Camacho-Ávila, A., Juárez-García, A. & Placencia-Reyes, O. (2015). Síndrome de Burnout en Instructores Comunitarios: Propiedades Psicométricas del Maslach Burnout Inventory General Survey (MBI-GS). En Juárez-García, A. (Ed.), *Evaluación Psicosocial en el Trabajo. Modelos, Escalas y Propiedades Psicométricas en Muestras Mexicanas* (pp. 279-308). México, DF: Plaza y Valdés.

Forero, C. G., Maydeu-Olivares, A., Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16, 625-641. <https://doi.org/10.1080/10705510903203573>

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50.
<https://doi.org/10.2307/315q1312>

- Franke, G. & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing: a comparison of four procedures. *Internet Research*, 29(3), 430-447. <https://doi.org/10.1108/IntR-12-2017-0515>
- Free, H., Groenewold, M. R., & Luckhaupt, S. E. (2020). Lifetime prevalence of self-reported work-related health problems among U.S. Workers - United States. *Morbidity and Mortality Weekly Report*, 69, 361–365. <https://doi.org/10.15585/mmwr.mm6913a1>
- Fuchs, C. & Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research: Conceptual issues and application guidelines. *Die Betriebswirtschaft; Stuttgart*, 69(2), 195-210. <https://doi.org/10.1007/s11747-011-0300-3>
- Fujita, A., Sato, J. R., Demasi, M. A., Sogayar, M. C., Ferreira, C. E., & Miyano, S. (2009). Comparing Pearson, Spearman and Hoeffding's D measure for gene expression association analysis. *Journal of Bioinformatics and Computational Biology*, 7(4), 663–684. <https://doi.org/10.1142/s0219720009004230>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 1, 156–168. <https://doi.org/10.1177/2515245919847202>
- Galtung, J. (1969). *Theory and Methods of Social Research*. Oslo: Universitetsforlaget.
- Ganster, D. C., & Rosen, C. C. (2013). Work stress and employee health: A multidisciplinary review. *Journal of Management*, 39(5), 1085–1122. <https://doi.org/10.1177/0149206313475815>
- García-Arroyo, J., & Osca, A. (2017). Coping with burnout: Analysis of linear, non-linear and interaction relationships. *Anales de Psicología*, 33(3), 722-731. <http://dx.doi.org/10.6018/analesps.33.3.279441>
- Gil-Monte, P. R. (2005). *El síndrome de quemarse por el trabajo (burnout): una enfermedad laboral en la sociedad del bienestar*. Madrid: Pirámide.

- Gil-Monte, P. R. (2009). Algunas razones para considerar los riesgos psicosociales en el trabajo y sus consecuencias en la salud pública. *Revista Española de Salud Pública*, 83(2), 169-73.
<https://doi.org/10.1590/s1135-57272009000200003>
- Gil-Monte, P. R. (2012). Riesgos psicosociales en el trabajo y salud ocupacional. *Revista Peruana de Medicina Experimental y Salud Pública*, 29(2), 237-341. <https://doi.org/10.1590/s1726-46342012000200012>
- Gil-Monte, P. R. (2016a). La Batería UNIPSICO: propiedades psicométricas de las escalas que evalúan los factores psicosociales de demanda. *Archivos de Prevención de Riesgos Laborales*, 19(2), 86-94. <https://dx.doi.org/10.12961/aprl.2016.19.02.3>
- Gil-Monte, P. R., López-Vílchez, J., Llorca-Rubio, J., & Sánchez-Piernas, J. (2016). Prevalencia de riesgos psicosociales en personal de la administración de justicia de la comunidad valenciana (España). *Liberabit*, 22, 7-19. <https://doi.org/10.24265/liberabit.2016.v22n1.01>
- Glass, G. V. (1966). Note on rank biserial correlation. *Educational and Psychological Measurement*, 26, 623-631. <https://doi.org/10.1177/001316446602600307>
- González-Trijueque, D., & Delgado Marina, S. (2013). Mobbing en trabajadores españoles y peruanos: un estudio exploratorio con el LIPT-60. *Liberabit*, 19(1), 121-131.
- Goodman, M. S., Griffey, R. T., Carpenter, C. R., Blanchard, M., & Kaphingst, K. A. (2015). Do subjective measures improve the ability to identify limited health literacy in a clinical setting? *The Journal of the American Board of Family Medicine*, 28(5), 584-594.
<https://doi.org/10.3122/jabfm.2015.05.150037>
- Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research Nursing Health*, 20(3), 269–274. [https://doi.org/10.1002/\(sici\)1098-240x](https://doi.org/10.1002/(sici)1098-240x)
- Gravetter, F., & Forzano, L. (2012). *Research Methods for the Behavioral Sciences*. Stamford, CT: Cengage Learning.

- Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, *74*(1), 155–167. <https://doi.org/10.1007/s11336-008-9099-3>
- Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, *74*, 121–135. <https://doi.org/10.1007/s11336-008-9098-4>
- Guidotti, T. L. (2000). Occupational epidemiology. *Occupational Medicine*, *50*(2), 141-145.
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinic Epidemiology*, *46*, 1417-1432.
- Guillot-Valdés, M., Guillén-Riquelme, A., & Buela-Casal, G. (2022). Content validity through expert judgment for the Depression Clinical Evaluation Test. *International Journal of Clinical and Health Psychology*, *22*, 1-8. <https://doi.org/10.1016/j.ijchp.2022.100292>
- Gulliksen, H. (1950). *Theory of mental tests*. John Wiley & Sons Inc.
- Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*, 29-48.
- Gwet, K. L. (2019). irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC). R package version 1.0. <https://CRAN.R-project.org/package=irrCAC>
- Hahr, Y., Reichenheim, M. E., Faerstein, E., Lambert, S., Fritzell, J., Toivanen, S., & Westerlund, H. (2014). Cross-cultural validity of the demand-control questionnaire: Swedish and Brazilian workers. *Revista de Saúde Pública*, *48*(3), 486-496. <https://doi.org/10.1590/S0034-8910.2014048005126>
- Hamui-Sutton, A. (2013). Un acercamiento a los métodos mixtos de investigación en educación médica. *Investigacion en Educación Médica*, *2*(8), 211-216. [https://doi.org/10.1016/S2007-5057\(13\)72714-5](https://doi.org/10.1016/S2007-5057(13)72714-5)

- Hancock G. R., Mueller R. O. (2000). Rethinking construct reliability within latent variable systems. In Cudek R., duToit S. H. C., Sorbom D. F. (Eds.), *Structural equation modeling: Present and future* (pp. 195-216). Lincolnwood, IL: Scientific Software.
- Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*, 373–388. <https://doi.org/10.1007/bf02294440>
- Harris, M. M., & Bladen, A. (1994). Wording effects in the measurement of role conflict and role ambiguity: A multitrait-multimethod análisis. *Journal of Management*, *20*(4), 887-901. [https://doi.org/10.1016/0149-2063\(94\)90034-5](https://doi.org/10.1016/0149-2063(94)90034-5)
- Harvey, S. B., Modini, M., Joyce, S., Milligan-Saville, J. S., Tan, L., Mykletun, A., Bryant, R. A., Christensen, H., & Mitchell, P. B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and Environmental Medicine*, *74*(4), 301–310. <https://doi.org/10.1136/oemed-2016-104015>
- Harvey, S. B., Modini, M., Joyce, S., Milligan-Saville, J. S., Tan, L., Mykletun, A., Bryant, R. A., Christensen, H., & Mitchell, P. B. (2017). Can work make you mentally ill? A systematic meta-review of work-related risk factors for common mental health problems. *Occupational and environmental medicine*, *74*(4), 301–310. <https://doi.org/10.1136/oemed-2016-104015>
- Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country Study. *International Journal of Cross Cultural Management*, *6*(2), 243 – 266. <https://doi.org/10.1177/1470595806066332>
- He, J., & Van de Vijver, F. J. R. (en prensa). Choosing an adequate design and analysis in cross-cultural personality research. *Current Issues in Personality Psychology*.
- Health & Safety Executive (s/f). *HSE Analysis Tool User Manual*. Disponible en: www.hse.gov.uk/stress/standards/pdfs/analysistoolmanual.

- Helson, R., & Soto, C. J. (2005). Up and down in middle age: Monotonic and nonmonotonic changes in roles, status, and personality. *Journal of Personality and Social Psychology*, 89(2), 194–204. <https://doi.org/10.1037/0022-3514.89.2.194>
- Hemker, B. T., Sijtsma, K., & Molenaar, I. W. (1995). Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken IRT model. *Applied Psychological Measurement*, 19(4), 337–352 <https://doi.org/10.1177/014662169501900404>
- Hemker, B. T., Sijtsma, K., Molenaar, I. W., & Junker, B. W. (1997). Stochastic ordering using the latent trait and the sum score in polytomous IRT models. *Psychometrika*, 62, 331-347. <https://doi.org/10.1007/BF02294555>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hernández, R. M., Cabrera-Orozco, I., Carranza, R., Mamani-Benito, O., & Turpo, O. (2021). Latin American scientific production on burnout in Scopus, 2010 - 2020. *Journal of Educational and Social Research*, 11(6), 186-15. <https://doi.org/10.36941/jesr-2021-0139>
- Hertzong, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health*, 31, 180-191.
- Hilkenmeier, F., Bohndick, C., Bohndick, T., & Hilkenmeier, J. (2020). Assessing distinctiveness in multidimensional instruments without access to raw data - a manifest fornell-larcker criterion. *Frontiers in psychology*, 11, 223. <https://doi.org/10.3389/fpsyg.2020.00223>
- Hindricks, I. (2020). Analizar con la ayuda de plantillas: la propuesta del Template Analysis. En: I., Hindricks. (Ed.), *Entre el campo y la Teoría. Estrategias de categorización inductiva y deductiva en investigación cualitativa*. México: Colegio de México.

- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236. <https://doi.org/10.1093/pan/mpl013>
- Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1-28. <https://doi.org/10.18637/jss.v042.i08>
- Hohwü, L., Lyshol, H., Gissler, M., Jonsson, S. H., Petzold, M., & Obel, C. (2013). Web-Based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. *Journal of Medical Internet Research, 15*(8), e173.e. <https://doi.org/10.2196/jmir.2595>
- Hökerberg, Y. H., Aguiar, O. B., Reichenheim, M., Faerstein, E., Valente, J. G., Fonseca, M., & Passos, S. R. (2010). Dimensional structure of the demand control support questionnaire: a Brazilian context. *International Archives of Occupational and Environmental Health, 83*(4), 407–416. <https://doi.org/10.1007/s00420-009-0488-4>
- Holbrook, A. (2008). Self-reported measure. In Lavrakas' (Ed.). *Encyclopedia of Survey Research Methods*. Thousand Oaks, CA: Sage. <http://dx.doi.org/10.4135/9781412963947.n523>
- Holtzman, N. S., & Donnellan, M. B. (2017). A simulator of the degree to which random responding leads to biases in the correlations between two individual differences. *Personality and Individual Differences, 114*, 187–192. <https://doi.org/10.1016/j.paid.2017.04.013>
- Horvat, M., & Tement, S. (2020). Self-reported cognitive difficulties and cognitive functioning in relation to emotional exhaustion: Evidence from two studies. *Stress and Health: Journal of the International Society for the Investigation of Stress, 36*(3), 350–364. <https://doi.org/10.1002/smi.2930>

- Houdmont, J., Jachens, L., Randall, R., Hopson, S., Nuttall, S., & Pamia, S. (2019). What Does a Single-Item Measure of Job Stressfulness Assess? *International journal of environmental research and public health*, 16(9), 1480. <https://doi.org/10.3390/ijerph16091480>
- Houdmont, J., Randall, R., Kerr, R., & Addley, K. (2013). Psychosocial risk assessment in organizations: Concurrent validity of the brief version of the Management Standards Indicator Tool. *Work & Stress*, 27(4), 403-412. <https://doi.org/10.1080/02678373.2013.843607>
- Houtman, I., Jettinghoff, K., & Cedillo, L. (2007). Raising awareness of stress at work in developing countries: a modern hazard in a traditional working environment (Protecting Workers' Health Series No 6). Geneva: World Health Organization: Geneva.
- Hoyle, R. H. & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 302-317). Thousand Oaks, CA: SAGE. <https://www.doi.org/10.4135/9781412986311>
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114. <https://doi.org/10.1007/s10869-011-9231-8>
- Hunsley, J. & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology*, 3, 29-51. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091419>
- Hurrell, J. J. Jr & Lindström, K. (1992). Comparison of job demands, control and psychosomatic complaints at different career stages of managers in Finland and the United States. *Scand J Work Environ Health* 18 Suppl. 2:11-13

- İnal, H., Koğar, E. Y., Demirdüzen, E., & Gelbal, S. (2017). Cronbach's coefficient alpha: A meta-analysis study. *H.U. Journal of Education*, 32(1), 18-32.
<http://doi.org/10.16986/HUJE.2016017219>
- Inceoglu, I., & Fleck, S. (2010). Engagement as a motivational construct. In S. L. Albrecht (Ed.), *Handbook of employee engagement: Perspectives, issues, research and practice* (pp. 74–86). Cheltenham Glos, UK: Edward Elgar Publishing.
<https://doi.org/10.4337/9781849806374.00012>
- International Labour Organisation (1986). *Psychosocial factors at work: recognition and control*. Report of the Joint ILO/ WHO Committee on Occupational Health. Ninth Session, Geneva, Setiembre, 18-24 (Occupational Safety and Health Series, 56). Geneva. Disponible en:
http://www.ilo.org/public/libdoc/ilo/1986/86B09_301_engl.pdf
- International Labour Organization (2016). *Workplace Stress: A collective challenge*. Turin, Italy: International Labour Organization. Disponible en:
https://www.ilo.org/safework/info/publications/WCMS_466547/lang--en/index.htm
- International Test Commission (2016). *The ITC Guidelines for Translating and Adapting Tests (Second edition)*. Disponible en: https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf
- Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema*, 26(3), 395-400. <https://doi.org/10.7334/psicothema2013.349>
- Jackson, D. L., Gillaspay, J. A. Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14, 6-23.
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods*, 14(1), 6–23.
<https://doi.org/10.1037/a0014694>

- Jacobo, G., & Máynez, A. (2020). Design and validity of a questionnaire for measuring psychosocial risk factors and burnout. *DYNA*, 87(214), 66-74. <https://doi.org/10.15446/dyna.v87n214.85187>
- Jacukowicz, A., & Wezyk, A. (2017). Development and validation of the Psychosocial Risks Questionnaire for Musicians (PRQM). *Psychology of Music*, 46(2), 252-265. <https://doi.org/10.1177/0305735617706540>
- Jang, E. E., & Roussos, L. (2007). An investigation into the dimensionality of TOEFL using conditional covariance-based nonparametric approach. *Journal of Educational Measurement*, 44(1), 1-21. <https://doi.org/10.1111/j.1745-3984.2007.00024.x>
- Jazari, M. D., Jahangiri, M., Khaleghi, H., Abbasi, N., Hassanipour, S., Shakerian, M., & Kamalinia, M. (2018). Prevalence of self-reported work-related illness and injuries among building construction workers, Shiraz, Iran. *EXCLI Journal*, 17, 724–733. <https://doi.org/10.17179/excli2018-1459>
- Jensen, I., & Harms-Ringdahl, K. (2007). Strategies for prevention and management of musculoskeletal conditions: neck pain. *Best Practice Research in Clinical Rheumatology*, 21, 93-108.
- Jimenez, P., & Dunkl, A. (2017). Assessment of psychosocial risks and mental stress at work: the development of the instrument OrgFit. *Journal of Ergonomics*, 7(2), 1. <https://doi.org/10.4172/2165-7556.1000188>
- Johnson, B. & Onwuegbuzie, A. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. <https://doi.org/10.3102/0013189x033007014>
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39(1), 103-129. <https://doi.org/10.1016/j.jrp.2004.09.009>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.

- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L.A. (2007). Toward a Definition of Mixed Methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36(2), 264-277. <https://doi.org/10.1177/0022022104272905>
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202. <https://doi.org/10.1007/BF02289343>
- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika* 57, 409-426
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). SemTools: Useful tools for structural equation modeling. R package version 0.5-5. <https://CRAN.R-project.org/package=semTools>
- Juárez-García, A. (2005). *Factores psicosociales y personalidad en relación a la salud laboral*. Tesis de doctorado. Universidad Nacional Autónoma de México.
- Juárez-García, A. (2007). Factores psicosociales, estrés y salud en distintas ocupaciones: un estudio exploratorio. *Investigación en salud*, 9(1), 57-64.
- Juárez-García, A. (2007). La dimensión de fatiga-energía como indicador de presentismo: validez de una escala en trabajadores mexicanos. *Ciencia y Trabajo*, 9(24), 55-60.
- Juárez-García, A. (2007). Psicología de la salud ocupacional: área de oportunidad en México. *Revista de Salud Pública y Nutrición*, 8(3), 1-2. Disponible en: <https://www.medigraphic.com/pdfs/revsalpubnut/spn-2007/spn073a.pdf>
- Juárez-García, A. (2017). ¿Qué son los factores psicosociales del trabajo? En: G. Blanco & L. Leyman (Eds.), *Factores psicosociales laborales y sus efectos, haciendo visible lo invisible* (pp. 13-42). Caracas: Araca Editores

- Juárez-García, A. (2017). *Proyecto: Observatorio de factores psicosociales y bienestar organizacional*. Autor: México.
- Juárez-García, A. (2018). Factores psicosociales del trabajo en México: Historia, conceptos y perspectivas. En H. Littlewood & F. Uribe, *Psicología organizacional en Latinoamérica* (pp. 89-108). México: UNAM y Manual Moderno.
- Juárez-García, A. (s/f). Clasificación de dimensiones de la encuesta de Procesos Psicosociales del Trabajo (PROPSIT). Documento no publicado: Autor.
- Juárez-García, A., & Andrade, P. (2004). Redes semánticas de trabajo, salud y relaciones interpersonales en el ámbito laboral de diferentes ocupaciones. *Revista de Psicología Social y Personalidad*, 20(1), 43-63.
- Juárez-García, A., & Camacho, A. (2011). Factores psicosociales de la salud en el trabajo: análisis de su concepción y base teórica. En A. Juárez y A. Camacho (Eds.), *Reflexiones teórico-conceptuales de lo psicosocial en el trabajo* (pp. 187-217). Cuernavaca, Morelos: Juan Pablos Editor y Universidad Autónoma del Estado de Morelos.
- Juárez-García, A., & Flores-Jiménez, C. (2020). Estructura factorial de un Instrumento para la Evaluación de Procesos Psicosociales en el Trabajo en México. *Revista de Psicología y Ciencias del Comportamiento de la Unidad Académica de Ciencias Jurídicas y Sociales*, 11(1), 181-202. <https://doi.org/10.29059/rpcc.20200617-111>
- Juárez-García, A., (2015). *Encuesta para evaluación de riesgos psicosociales (PROPSIT)*. Autor: México.
- Juárez-García, A., Camacho-Ávila, A., García-Rivas, J., & Gutiérrez-Ramos O. (2021). Psychosocial factors and mental health in Mexican healthcare workers during the COVID-19 pandemic. *Salud Mental*, 44(5), 229-240. <https://doi.org/10.17711/SM.0185-3325.2021.030>

- Juárez-García, A., Flores, J. C., & Imke, H. (2019). Exploración de factores psicosociales del trabajo mediante una técnica de análisis mixto: La transformación de categorías en variables. En: I. Hindrichs (Ed.), *Entre el campo y la teoría* (pp. 241-275.). México: Colegio Mexiquense
- Juárez-García, A., Flores-Jiménez, C. A., & Pelcastre-Villafuerte, B. E. (2020). Factores psicosociales del trabajo y efectos psicológicos en comerciantes informales en Morelos, México: Una exploración mixta preliminar. *Salud UIS*, 52(4), 402-413.
<https://doi.org/10.18273/revsal.v52n4-2020007>
- Juárez-García, A., Flores-Jiménez, C., & Pelcastre-Villafuerte, B. (2020). Factores psicosociales del trabajo y efectos psicológicos en comerciantes informales en Morelos, México: una exploración mixta preliminar. *Revista de la Universidad Industrial de Santander. Salud*, 52(4), 402-413.
<https://doi.org/10.18273/revsal.v52n4-2020007>
- Juárez-García, A., García, J., & Vera, A. (2010). Occupational Health Psychology in Latin-America: The Networking Efforts. *Society for Occupational Health Psychology Newsletter*, 8, 2-4.
- Juárez-García, A., Merino Soto, C., Fernández Arata, M., Flores Jiménez, C. A., Caraballo, M., & Camacho Cristiá, C. (2020). Validación transcultural y funcionamiento diferencial del MBI-GS en docentes de tres países latinoamericanos. *Avances en Psicología Latinoamericana*, 38(1), 135-156. <https://doi.org/10.12804/revistas.urosario.edu.co/apl/a.6621>
- Juárez-García, A., Vera, A., Merino, C., Gómez, V., Feldman, L., & Hernández, E. (2014). Demanda/Control y la Salud Mental en Profesionales de la Salud: Un Estudio en Seis Países Latinoamericanos. *Informació Psicológica*, 108, 2-18.
- Kalimo, R., El Batawi, M. A. & Cooper, C. L. (1987). *Psychosocial factors at work and their relation to health*. Geneva: World Health Organization.

- Kam C. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in psychology, 10*, 1258.
<https://doi.org/10.3389/fpsyg.2019.01258>
- Kamata, A. & Bauer, D. J. (2008) A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling, 15*(1) 136-153.
<https://doi.org/10.1080/10705510701758406>
- Kane, M., & Bridgeman, B. (2017). Research on validity theory and practice at ETS. In R. E. Bennett & M. von Davier (Eds.), *Methodology of educational measurement and assessment. Advancing human assessment: The methodological, psychological and policy contributions of ETS* (p. 489–552). New York, US: Springer Science + Business Media. https://doi.org/10.1007/978-3-319-58689-2_16
- Karasek, R. & Theorell, T. (1990). *Healthy work: stress, productivity, and the reconstruction of working life*. New York: Basic Books.
- Karasek, R. A. (1998). Demand/Control Model: A social, emotional, and physiological approach to stress risk and active behaviour development. In J. M. Stellman (Ed.), *Encyclopaedia of occupational health and safety* (pp. 34.6-34.14). Geneva: ILO.
- Karasek, R. A., & Theorell, T. (1990). *Healthy work: Stress, productivity and the reconstruction of working life*. Nueva York: Basic Books.
- Karlqvist, L., Wigaeus Tornqvist, E., Hagberg, M., Hagman, M. T. & oomingas, A. (2002). Self-reported working conditions of VDU operators and associations with musculoskeletal symptoms: a cross-sectional study focusing on gender differences. *International Journal of Industrial Ergonomics, 30*, 4-5, 277-294. [https://doi.org/10.1016/s0169-8141\(02\)00131-2](https://doi.org/10.1016/s0169-8141(02)00131-2)
- Kelley, K. (2020). MBESS: The MBESS R Package. R package version 4.8.0. <https://CRAN.R-project.org/package=MBESS>

- Keselman, H. J., Othman, A. R., & Wilcox, R. R. (2013). Preliminary testing for normality: is this a good practice? *Journal of Modern Applied Statistical Methods*, 12(2), 2.
<https://doi.org/10.22237/jmasm/1383278460>
- Khamis, H. (2008). Measures of association: How to choose? *Journal of Diagnostic Medical Sonography*, 24(3), 155–162. <https://doi.org/10.1177/8756479308317006>
- Kim, H. (1994). New techniques for the dimensionality assessment of standardized test data (Doctoral dissertation, University of Illinois at Urbana-Champaign). <http://hdl.handle.net/2142/19110>
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press,
- Knekta, E., Runyon, C., & Eddy, S. (2019). One size doesn't fit all: Using factor analysis to gather validity evidence when using surveys in your research. *CBE life sciences education*, 18(1), rm1.
<https://doi.org/10.1187/cbe.18-04-0064>
- Kořar, H. (2018). An examination of parametric and nonparametric dimensionality assessment methods with exploratory and confirmatory modeling. *Journal of Education and Learning*, 7(3), 148-158.
- Korkmaz, S., Goksuluk, D., Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, 6(2), 151-162. <https://doi.org/10.32614/RJ-2014-031>
- Kristensen, T.S., Hannerz, H., Høgh, A., & Borg, V. (2005). The Copenhagen Psychosocial Questionnaire—a tool for the assessment and improvement of the psychosocial work environment. *Scandinavian Journal of Work Environment & Health*, 31(6), 438–449.
<https://doi.org/10.5271/sjweh.948>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical care*, 41(11), 1284–1292.
<https://doi.org/10.1097/01.MLR.0000093487.78664.3C>

- Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2009). An ultra-brief screening scale for anxiety and depression: The PHQ-4. *Psychosomatics*, 50(6), 613-621.
[https://doi.org/10.1016/S0033-3182\(09\)70864-3](https://doi.org/10.1016/S0033-3182(09)70864-3)
- Kuhn, T. (1962). *La estructura de las revoluciones científicas* [1ra ed. Español: 1971]. México, DF: Fondo de Cultura Económica.
- Kurtz, J. E., & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment*, 76(2), 315-332. https://doi.org/10.1207/S15327752JPA7602_12
- Kvale, S. (2011). *Las entrevistas en investigación cualitativa*. Madrid: Ediciones Morata.
- Lahner, F. M., Schauber, S., Lörwald, A. C., Kropf, R., Guttormsen, S., Fischer, M. R., & Huwendiek, S. (2020). Measurement precision at the cut score in medical multiple choice exams: Theory matters. *Perspectives on Medical Education*, 9(4), 220–228. <https://doi.org/10.1007/s40037-020-00586-0>
- Landsbergis, P., Theorell, T., Schwartz, J., Greiner, B. A., & Krause, N. (2000). Measurement of psychosocial workplace exposure variables. *Occupational Medicine*, 15(1), 163-188.
- Lavelle, E., Vuk, J., & Barber, C. (2013). Twelve tips for getting started using mixed methods in medical education research. *Medical Teaching*, 35, 272-276.
<https://doi.org/10.3109/0142159x.2013.759645>
- Law, G. (2011). Error in the fernandez huerta readability formula. Disponible en:
<https://linguistlist.org/issues/22/22-2332.html>
- Lázarus, S. & Folkman, S. (1986). *Estrés y procesos cognitivos*. Barcelona: Martínez Roca.
- Leech, N. L., & Onwuegbuzie, A. J. (2009). A Typology of Mixed Methods Research Designs. *Quality & Quantity*, 43(2), 265-275. <https://doi.org/10.1007/s11135-007-9105-3>

- Leibler, J. H., & Perry, M. J. (2017). Self-reported occupational injuries among industrial beef slaughterhouse workers in the Midwestern United States. *Journal of Occupational and Environmental Hygiene*, 14(1), 23–30. <https://doi.org/10.1080/15459624.2016.1211283>
- Leka, S., Jain, A. & World Health Organization (2010). Health impact of psychosocial hazards at work: an overview. Geneva: World Health Organization.
- LePine, J. A., LePine, M. A., & Jackson, C. L. (2004). Challenge and hindrance stress: relationships with exhaustion, motivation to learn, and learning performance. *Journal of Applied Psychology*, 89(5), 883-891. doi: <https://doi.org/10.1037/0021-9010.89.5.883>.
- Li, Ch. (2016a). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-49. <https://doi.org/10.3758/s13428-015-0619-7>
- Li, Ch. (2016b). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, 21(3), 369-87. <https://doi.org/10.1037/met0000093>
- Liao, P., & Hsieh, J. (2017). Does internet-based survey have more stable and unbiased results than paper-and-pencil survey? *Open Journal of Social Sciences*, 5, 69-86. <https://doi.org/10.4236/jss.2017.51006>
- Liao, T. F. (2002). *Statistical group comparison*. New York: Wiley-InterScience.
- Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 2, 103-122.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.

- Lindley, P.A., Bartram, D., & Kennedy, N. (2008). *EFPA Review Model for the description and evaluation of psychological tests: Test review form and notes for reviewers: Version 3.42*. Brussels: EFPA Standing Committee on Tests and Testing (September, 2008).
- Lindstrom K., Elo A., Skogstad A., Dallner M., Gamberale F., Hottinen V., Knardahl S., & Orhede E. (2000). *User's Guide for the QPS Nordic – General Nordic Questionnaire for Psychological and Social Factors at Work*. Copenhagen: Nordic Council of Ministers.
- Lindström, K. & Hurrell, J. J. Jr. (1992). Coping with job stress by managers at different career stages in Finland and the United States. *Scandinavian Journal of Work Environment Health, 18* (Suppl. 2), 14-17.
- Lindström, K. (2011). Career Stage. In Lindstrom, C. Lindstrom, C., J. Levi, L. R. Murphy, S. L. Sauter (Eds.). *Encyclopaedia of Occupational Health and Safety* (Vol. 3., pp. 41-42). Geneva: International Labour Organization.
- Linton, S. J. & van Tulder, M.W. (2001). Preventive interventions for back and neck pain problems: what is the evidence? *Spine, 26, 7, 778-787*. <https://doi.org/10.1097/00007632-200104010-00019>
- Littman, A. J., White, E., Satia, J. A., Bowen, D. J., & Kristal, A. R. (2006). Reliability and validity of 2 single-item measures of psychosocial stress. *Epidemiology, 17*(4), 398-403. <https://doi.org/10.1097/01.ede.0000219721.89552.51>
- Lloret, S., Ferreres, A., Hernández, A. y Tomás, I. (2014). El análisis factorial exploratorio de los ítems: Una guía práctica, revisada y actualizada. *Anales de Psicología, 30*(3), 1151-1169.
- Lloret, S., Ferreres, A., Hernández, A., & Tomás, I. (2017). The exploratory factor analysis of items: guided analysis based on empirical data and software. *Anales de Psicología, 33*(2), 417-432. <https://doi.org/10.6018/analesps.33.2.270211>

- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694. <http://dx.doi.org/10.2466/PR0.3.7.635-694>
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(2), 57–64. <https://doi.org/10.1027/1614-2241.2.2.57>
- Lu, M.-L., Nakata, A., Park, J. B., & Swanson, N. G. (2014). Workplace psychosocial factors associated with work-related injury absence: a study from a nationally representative sample of korean workers. *International Journal of Behavioral Medicine*, 21(1), 42–52. <http://doi.org/10.1007/s12529-013-9325-y>
- Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). parameters: Extracting, computing and exploring the parameters of statistical models using R. *Journal of Open Source Software*, 5(53), 2445. <https://doi.org/10.21105/joss.02445>
- Luethi, M., Meier, B., & Sandi, C. (2009). Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. *Frontiers in Behavioral Neuroscience*, 2, 5. <https://doi.org/10.3389/neuro.08.005.2008>
- Macias, C., Gold, P. B., Öngür, D., Cohen, B. M., & Panch, T. (2015). Are single-item global ratings useful for assessing health status? *Journal of Clinical Psychology in Medical Settings*, 22(4), 251-264. <https://doi.org/10.1007/s10880-015-9436-5>
- Mäder, I. A., & Niessen, C. (2017). Nonlinear associations between job insecurity and adaptive performance: The mediating role of negative affect and negative work reflection. *Human Performance*, 30(5), 231–253. <https://doi.org/10.1080/08959285.2017.1364243>
- Magnavita N. (2008). Strumenti per la valutazione dei rischi psicosociali sul lavoro [Questionnaires for psychosocial risk assessment at work]. *Giornale italiano di medicina del lavoro ed ergonomia*, 30(1 Suppl A), A87–A97.

- Magnavita, N., & Fileni, A. (2014). Association of work-related stress with depression and anxiety in radiologists. *La Radiología Médica*, *119*(5), 359–366. <https://doi.org/10.1007/s11547-013-0355-y>
- Makonnen, S. M. Y. (2019). A simulation study on the power of mutual information and distance correlation. Unpublished master's thesis psychology, Faculty of Social and Behavioral Sciences, Leiden University. Disponible en: <https://core.ac.uk/download/pdf/287722588.pdf>
- Mangiafico, S. (2020). *rcompanion*: functions to support extension education program evaluation. R package versión 2.3.25. <https://CRAN.R-project.org/package=rcompanion>
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61-83. <https://doi.org/10.1016/j.jrp.2013.09.008>
- Manzanares-Medina, E., Merino-Soto, C., & Fernández-Arata, M. (2016). Estructura interna del Maslach Burnout Inventory (MBI) en una muestra de sacerdotes y religiosas católicas peruanos. *Salud & Sociedad*, *7*(2), 198-211
- Marcatto, F., Colautti, L., Larese Filon, F., Luis, O., & Ferrante, D. (2014). The HSE Management Standards Indicator Tool: concurrent and construct validity. *Occupational Medicine (Oxford, England)*, *64*(5), 365–371. <https://doi.org/10.1093/occmed/kqu038>
- Marcum, J. I., Chin, B., Anderson, N. J., & Bonauto, D. K. (2017). Self-reported work-related injury or illness — Washington, 2011–2014. *Centers for Disease Control. Morbidity and Mortality Weekly Report*, *66*(11), 302–306. <https://doi.org/10.15585/mmwr.mm6611a6>
- Márquez, A., & Juárez, A. (2018). Efecto moderador de la abnegación en la relación hostigamiento laboral y malestar mental en trabajadores del sector público. *Enseñanza e Investigación en Psicología*, *23*(3), 292-302.

- Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1
- Martínez-García, J. & Martínez-Caro, L. (2009). La validez discriminante como criterio de evaluación de escalas: ¿teoría o estadística? *Universitas Psychologica*, 8(1), 27-36.
- Martinez-Gomez, M., Marin-Garcia, J., Giraldo-O'meara, M. (2017). Testing invariance between web and paper students satisfaction surveys: A case study. *Intangible Capital*, 13(5), 879-901. doi: <http://dx.doi.org/10.3926/ic.1049>.
- Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). *Maslach Burnout Inventory Manual (3a ed.)*. Palo Alto, CA: Consulting Psychologists Press.
- Masuda, S., Kitaoka, K., & Ogino, K., (2012). Effects of item wording direction and grouping on psychological measures. *The Japanese Journal of Health Psychology*, 25(1), 31-41. https://doi.org/10.11560/jahp.25.1_31
- Maydeu-Olivares, A. (2017). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, 24(3), 383–394. <https://doi.org/10.1080/10705511.2016.1269606>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: L. Erlbaum Associates.
- McDonald, R. P. (2005). Semiconfirmatory factor analysis: The example of anxiety and depression. *Structural Equation Modeling*, 12(1), 163-172. https://doi.org/10.1207/s15328007sem1201_9
- McDowell, J., Courtney, M., Edwards, H., & Shortridge-Baggett, L. (2005). Validation of the Australian/English version of the Diabetes Management Self-Efficacy Scale. *International Journal of Nursing Practice*, 11, 177-184. <https://doi.org/10.1111/j.1440-172X.2005.00518.x>

- McGonagle, A. K., Huang, J. L., & Walsh, B. M. (2016). Insufficient effort survey responding: An under-appreciated problem in work and organizational health psychology research. *Applied Psychology, 65*(2), 287–321. <https://doi.org/10.1111/apps.12058>
- McMurray, J. E., Linzer, M., Konrad, T. R., Douglas, J., Shugerman, R., & Nelson, K. (2000). The work lives of women physicians results from the physician work life study: The SGIM career satisfaction study group. *Journal of General Intern Medicine, 15*(6), 372–380. <https://doi.org/10.1111/j.1525-1497.2000.im9908009.x>
- McNaught, C., & Lam, P. (2010). Using Wordle as a Supplementary Research Tool. *The Qualitative Report, 15*(3), 630-643. <https://doi.org/10.46743/2160-3715/2010.1167>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437-455. <https://doi.org/10.1037/a0028085>
- Meijer, R. R., & Baneke, J. J. (2004). Analyzing psychopathology items: A case for nonparametric item response theory modeling. *Psychological Methods, 9*(3), 354–368. <https://doi.org/10.1037/1082-989X.9.3.354>
- Meijer, R. R., Sijtsma, K., & Smid, N. G. (1990). Theoretical and empirical comparison of the Mokken and the Rasch approach to IRT. *Applied Psychological Measurement, 14*, 283-298. <https://doi.org/10.1177/014662169001400306>
- Meijer, R. R., Tendeiro, J. N., & Wanders, R. B. K. (2015). *The use of nonparametric item response theory to explore data quality*. In S. P. Reise & D. A. Revicki (Eds.), *Multivariate applications series. Handbook of item response theory modeling: Applications to typical performance assessment* (p. 85–110). London, England: Routledge/Taylor & Francis Group.
- Menard, J. C., Hinds, P. S., Jacobs, S. S., Cranston, K., Wang, J., DeWalt, D. A., & Gross, H. E. (2014). Feasibility and acceptability of the PROMIS Measures in children and adolescents in

active cancer treatment and survivorship. *Cancer Nursing*, 37(1), 66-74.

<https://doi.org/10.1097/ncc.0b013e3182a0e23d>

Mercer, C. D. (1991a): *Dificultades de aprendizaje. 1. Origen y diagnóstico*. Barcelona: CEAC.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical care*, 44(11 Suppl 3), S69–S77. <https://doi.org/10.1097/01.mlr.0000245438.73837.89>

Merino Soto, C., Calderón-De la Cruz, G., Gil-Monte, P., & Arturo, J.-G. (2021). Validez sustantiva en el marco de la validez de contenido: Aplicación en la escala de Carga de Trabajo. *Revista Argentina De Ciencias Del Comportamiento*, 13(1), 81–92.

<https://doi.org/10.32348/1852.4206.v13.n1.20547>

Merino, C. (2011a). Datos normativos del desarrollo psicomotor: Un reanálisis. *Revista Peruana de Psicología*, 1(1), 30-41.

Merino-Soto, C. & Angulo-Ramos, M. (2020). Validity induction: comments on the study of Compliance Questionnaire for Rheumatology. *Revista Colombiana de Reumatología*. Advance online publication. <https://doi.org/10.1016/j.rcreu.2020.05.005>

Merino-Soto, C. & Calderón-De la Cruz, G. A. (2018). Validez de estudios peruanos sobre estrés y burnout. *Revista Peruana de Medicina Experimental y Salud Publica*, 35, 353-354.

<http://doi.org/10.17843/rpmesp.2018.352.3521>

Merino-Soto, C. & Juárez-García, A. (2019). Revisión sistemática de las publicaciones peruanas en psicología organizacional. Datos no publicados.

Merino-Soto, C. & Livia, C. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales en Psicología*, 25(1), 169-171

- Merino-Soto, C. (2016). Percepción de la claridad de los ítems: Comparación del juicio de estudiantes y jueces-expertos. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 14(2), 1469-1477.
- Merino-Soto, C. (2018). Confidence interval for difference between coefficients of content validity (Aiken's V): a SPSS syntax. *Anales de Psicología*, 34(3), 587-590.
<https://dx.doi.org/10.6018/analesps.34.3.283481>
- Merino-Soto, C. (2018). El participante sensible: extendiendo la metodología de la validez de contenido. *Revista Médica de Chile*, 146(12), 1497-1498. <https://dx.doi.org/10.4067/s0034-98872018001201497>
- Merino-Soto, C. (2018). El participante sensible: extendiendo la metodología de la validez de contenido. *Revista Médica de Chile*, 146(12), 1497-1498. <https://dx.doi.org/10.4067/s0034-98872018001201497>
- Merino-Soto, C., & Angulo-Ramos, M. (2021). Metric Studies of the Compliance Questionnaire on Rheumatology (CQR): A Case of Validity Induction? *Reumatología clinica*, S1699-258X(21)00093-0. <https://doi.org/10.1016/j.reuma.2021.03.004>
- Merino-Soto, C., & Livia, C. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: Un programa Visual Basic para la V de Aiken. *Anales en Psicología*, 25(1), 169-171.
- Merino-Soto, C., & Pérez, L. (En revisión). La reflexión en la investigación cuantitativa y el horizonte teórico-metodológico del investigador.
- Merino-Soto, C., Angulo-Ramos, M., & López-Fernández, V. (2019). Escala de inteligencia Emocional Wong-Law (WLEIS) en estudiantes de Enfermería peruanos. *Educación Médica Superior*, 33(1), e1473.

- Merino-Soto, C., Angulo-Ramos, M., & López-Fernández, V. (2019). Escala de inteligencia Emocional Wong-Law (WLEIS) en estudiantes de Enfermería peruanos. *Educación Médica Superior*, 33(1), e1473.
- Merino-Soto, C., Domínguez-Lara, S., & Fernández-Arata, M. (2017). Validación inicial de una Escala Breve de Satisfacción con los Estudios en estudiantes universitarios de Lima. *Educación Médica*, 18(1), 74-77. <https://doi.org/10.1016/j.edumed.2016.06.016>.
- Merino-Soto, C., Juárez-García, A., & Angulo-Ramos, M. (2017). Determinación de los niveles de burnout en médicos internos: cuestiones críticas. *Investigación en Educación Médica*, 6(24), 282. <https://doi.org/10.1016/j.riem.2016.09.004>
- Merino-Soto, C., Juarez-Garcia, A., Altamirano, A., & Velarde, B. (2018). Una medida muy breve del burnout: evidencia de validez de constructo en trabajadores peruanos. *Ansiedad y Estrés*, 24(2), 131-135. <http://doi.org/10.1016/j.anyes.2018.07.001>
- Merino-Soto, C., Lima-Mendoza, S., Lozano-Huamán, M., & Calderón – de la Cruz, G., & Juárez-García, Arturo. (2021). Escala de Auto-eficacia Ocupacional - Breve (OSES): exploración de sus propiedades psicométricas. *Revista de la Asociación Española de Especialistas en Medicina del Trabajo*, 30(2), 195-207.
- Merino-Soto, C., Lozano-Huamán, M., Lima-Mendoza, S., Calderón de la Cruz, G., Juárez-García, A., & Toledano-Toledano, F. (2022). Ultrashort Version of the Utrecht Work Engagement Scale (UWES-3): A Psychometric Assessment. *International Journal of Environmental Research and Public Health*, 19(2), 890. <https://doi.org/10.3390/ijerph19020890>
- Merletti, F., Soskolne, C. L., & Vineis, P. (1998). Epidemiological method applied to occupational health and safety. En: J.M. Stellman (Ed.), *Encyclopedia of Occupational Health and Safety* (4th. Ed., pp. 28.2-28.6). Genova: International Labour Office.

- Messick, S. (1989). Validity. En R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1994). Foundations of validity: Meaning and consequences in psychological assessment. *European Journal of Psychological Assessment*, *10*(1), 1-9. <https://doi.org/10.1002/j.2333-8504.1993.tb01562.x>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749. <https://doi.org/10.1037//0003-066x.50.9.741>
- Meyer, J. P. (2014). *Applied Measurement with jMetrik*. New York, NY: Routledge.
- Meyer, R. D., Mumford, T. V., Burrus, C. J., Campion, M. A., & James, L. R. (2014). Selecting null distributions when calculating rwg: a tutorial and review. *Organizational Research Methods*, *17*(3), 324–345. <https://doi.org/10.1177/1094428114526927>
- Meyerson, P., & Tryon, W.W. (2003) Validating internet research: a test of the psychometric equivalence of internet and in-person samples. *Behavior Research Methods, Instruments, and Computers*, *35*, 614-620. <https://doi.org/10.3758/BF03195541>
- Micceri T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105* 156–166. <https://doi.org/10.1037/0033-2909.105.1.156>
- Miecznikowski, J. C., Hsu, E-S., Chen, Y., & Vexler, A. (2017). testforDEP: Dependence tests for two variables. R package version 0.2.0. <https://CRAN.R-project.org/package=testforDEP>
- Ministerio de Salud (2005). *Manual de Salud Ocupacional*. Lima: Autor.
- Ministerio de Trabajo y Promoción del Empleo (2012). Reglamento de la Ley Nro 29783, Ley de Seguridad y Salud en el Trabajo. Lima: Autor.

- Mokkan, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, 6(4), 417–430.
<https://doi.org/10.1177/014662168200600404>
- Mokken, R. J. (1971). *A theory and procedure of scale analysis: with applications in political research* (Vol. 1). Mouton, Netherlands: Walter de Gruyter.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). 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*, 19(4), 539-549. <https://doi.org/10.1007/s11136-010-9606-8>
- Mollenkopf, W.G. (1949). Variation of the standard error of measurement. *Psychometrika*, 14, 189–229. <https://doi.org/10.1007/BF02289153>
- Mommersteeg, P.M.C., Widdershoven, J.W.M.G., & Kop, W.J. (2017). Sex and gender differences in psychosocial risk factors for ischemic heart disease. In A.H.E.M. Maas & C.N.B. Merz (Eds.), *Manual of gynecardiology: Female-specific cardiology* (pp. 203-220). Springer International.
https://doi.org/10.1007/978-3-319-54960-6_13
- Monahan, P. O., Stump, T. E., Finch, H., & Hambleton, R. K. (2007). Bias of exploratory and cross-validated DETECT index under unidimensionality. *Applied Psychological Measurement*, 31(6), 483–503. <https://doi.org/10.1177/0146621606292216>
- Moncada, S., Llorens, C., & Kristensen, T. S. (2004). *Método ISTAS21 (CoPsoQ): Manual para la evaluación de riesgos psicosociales en el trabajo*. Madrid: Autor. Disponible en:
http://www.istas.ccoo.es/descargas/m_metodo_istas21.pdf
- Moncada, S., Llorens, C., Navarro, A. & Kristensen, T. S. (2005). ISTAS21: Versión en lengua castellana del cuestionario psicosocial de Copenhagen (COPSOQ). *Archivos de Prevención de Riesgos Laborales*, 8, 18-29.

- Morata-Ramirez, M. A., Holgado-Tello, F. P., Barbero-Garcia, I., & Mendez, G. (2015). Confirmatory factor analysis: Recommendations for unweighted least squares method related to Chi-Square and RMSEA Type I error. *Acción Psicológica*, *12*(1), 79-90.
<https://doi.org/10.5944/ap.12.1.14362>
- Morrison, K. W., & Embretson, S. (2018). Item Generation. In: P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development* (Vol. 1, pp. 74-94). Hoboken, NJ: John Wiley & Sons.
- Mullinix, K., Leeper, T., Druckman, J., & Freese, J. (2015). The generalizability of survey experiments. *Journal of Experimental Political Science*, *2*(2), 109-138.
<http://doi.org/10.1017/XPS.2015.19>
- Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición. *Psicothema*, *25*(2), 151-157.
- Muñoz, A. (2020). Analizador de legibilidad de texto. Disponible en: <https://legible.es/>
- Muñoz, M. (2006). Legibilidad y variabilidad de los textos. *Boletín de Investigación Educativa*, *Pontificia Universidad Católica de Chile*, *21*(2), 13-26.
- Muthén, B. O. (1993). Goodness of fit with categorical and nonnormal variables. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 205-234). Newbury Park, CA: Sage.
- Muthén, B. O., du Toit, S. H. C., Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Disponible en: https://www.statmodel.com/download/Article_075.pdf
- Nahum-Shani, I., Bamberger, P. A., & Bacharach, S. B. (2011). Social support and employee well-being: the conditioning effect of perceived patterns of supportive exchange. *Journal of Health and Social Behavior*, *52*(1), 123–139. <https://doi.org/10.1177/0022146510395024>

- Nandakumar, N. (2016). Assessing essential unidimensionality of real data. *Applied Psychological Measurement, 17*(1), 29-38. <https://doi.org/10.1177/014662169301700108>
- Nandakumar, R. (1991), Traditional dimensionality versus essential dimensionality. *Journal of Educational Measurement, 28*, 99-117. <https://doi.org/10.1111/j.1745-3984.1991.tb00347.x>
- National Institute for Occupational Safety and Health. (1999). STRESS...At Work (NIOSH Publication No. 99-101). Recuperado de: <http://www.cdc.gov/niosh/docs/99-101/>
- Netherlands Cancer Institute, Amsterdam. (2002). Assessing health status and quality-of-life instruments: attributes and review criteria. *Quality of Life Research, 11*(3), 193-205.
- Ng, C., & Chan, V. (2021). Prevalence of workplace bullying and risk groups in chinese employees in hong kong. *International Journal of Environmental Research and Public Health, 18*(1), 329. <https://doi.org/10.3390/ijerph18010329>
- Niblo, D. M. & Jackson, M. S. (2004). Model for combining the qualitative emic approach with the quantitative derived etic approach. *Australian Psychologist, 39*(2), 127 – 133. <https://doi.org/10.1080/00050060410001701843>
- Nielsen, K., & Mikkelsen, K. (2007). Predictive factors for self-reported occupational injuries at 3 manufacturing plants. *Safety Science Monitor, 11*(2), 1–9.
- Nieto-Gutierrez, W., Toro-Huamanchumo, C. J., Taype-Rondan, A., Timaná-Ruiz, R., Alva Diaz, C., Jumpa-Armas, D., Escobedo-Palza, S., & CONAREME Consejo Nacional de Residentado Médico (2018). Workplace violence by specialty among Peruvian medical residents. *PloS One, 13*(11), e0207769. <https://doi.org/10.1371/journal.pone.0207769>
- Nieuwenhuijsen, K., Bruinvels, D., & Frings-Dresen, M. (2010). Psychosocial work environment and stress-related disorders, a systematic review. *Occupational medicine (Oxford, England), 60*(4), 277–286. <https://doi.org/10.1093/occmed/kqq081>

- Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology, 96*(5), 966-980. <https://doi.org/10.1037/a0022955>
- Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2019). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. *Organizational Research Methods, 22*(3), 678-709. <https://doi.org/10.1177/1094428118761122>
- Olmos, A., & Govindasamy, P. (2015). Propensity scores: A practical introduction using R. *Journal of MultiDisciplinary Evaluation, 11*(25), 68-88.
- Olson, K., Smyth, J. D., Wang, Y., & Pearson, J. E. (2011). The self-assessed literacy index: Reliability and validity. *Social Science Research, 40*(5), 1465–1476.
<https://doi.org/10.1016/j.ssresearch.2011.05.002>
- Omholt, M. L., Tveito, T. H., & Ihlebæk, C. (2017). Subjective health complaints, work-related stress and self-efficacy in Norwegian aircrew. *Occupational medicine (Oxford, England), 67*(2), 135–142. <https://doi.org/10.1093/occmed/kqw127>
- Onwuegbuzie, A. J., & Johnson, R. B. (2006). The validity issue in mixed research. *Research in the Schools, 13*, 48-63.
- Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research, 4*, 56 – 78.
<https://doi.org/10.1177/1558689809355805>
- Onwuegbuzie, A.J., Daniel, L.G., & Collins, K. M. T. (2009). A meta-validation model for assessing the score validity of student teacher evaluations. *Quality & Quantity: International Journal of Methodology, 43*, 197-209. <https://doi.org/10.1007/s11135-007-9112-4>
- Onyishi, I. E., Ugwu, F. O., Onyishi, C. N., & Okwueze, F. O. (2018). Job demands and psychological well-being: Moderating role of occupational self-efficacy and job social support among mid-

career academics. *Journal of Psychology in Africa*, 28(4), 267–272.

<https://doi.org/10.1080/14330237.2018.1501908>

Oppenheimer, D. M., Meyvis, T., Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867-872. <https://doi.org/10.1016/j.jesp.2009.03.009>

Organización Internacional del Trabajo / Organización Mundial de la Salud (1984). *Factores psicosociales en el trabajo: Naturaleza, incidencia y prevención*. Informe del comité mixto OIT/OMS sobre medicina del trabajo. Ginebra, Suiza: OIT/OMS.

Orhede, E., & Kreiner, S. (2000). Item bias in indices measuring psychosocial work environment and health. *Scandinavian Journal of Work, Environment & Health*, 26(3), 263–272.

<https://doi.org/10.5271/sjweh.541>

Osborne, J. W. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research & Evaluation*, 20(2). <http://pareonline.net/getvn.asp?v=20&n=2>

Osterling, J. (2008). Psychometric properties. En: F. T. L. Long (Ed.) *Encyclopedia of counseling* (Vol. 2, pp. 804-808). Thousand Oaks, California: Sage.

Ostry, A. S., Kelly, S., Demers, P. A., Mustard, C., & Hertzman, C. (2003). A comparison between the effort-reward imbalance and demand control models. *BMC Public Health*, 3, 10.

<http://doi.org/10.1186/1471-2458-3-10>

Padilla, M. A., & Divers, J. (2013a). Bootstrap interval estimation of reliability via coefficient omega. *Journal of Modern Applied Statistical Methods*, 12(1), 78-89.

<https://doi.org/10.22237/jmasm/1367381520>

Padilla, M. A., & Divers, J. (2013b). Coefficient omega bootstrap confidence intervals: Nonnormal distributions. *Educational and Psychological Measurement*, 73(6), 956–972.

<https://doi.org/10.1177/0013164413492765>

- Pando-Moreno, M., Varillas, W., Aranda-Beltrán, C., & Elizalde-Núñez, F. (2016). Análisis factorial exploratorio del 'Cuestionario de factores psicosociales en el trabajo' en Perú. *Anales de la Facultad de Medicina*, 77(4), 365-371. <http://dx.doi.org/10.15381/anales.v77i4.12649>
- Parker, R. I., Hasbrouck, J. E., & Weaver, L. (2001). Spanish readability formulas for elementary-level texts: a validation study. *Reading & Writing Quarterly*, 17(4), 307-322. <http://dx.doi.org/10.1080/105735601317095052>
- Patlán, J. (2017). Salud ocupacional y psicología: ¿Quo vadis? *Revista Cubana de Salud y Trabajo*, 18(1), 67-73.
- Patton, M. (1990). *Qualitative evaluation and research methods* (pp. 169-186). Beverly Hills, CA: Sage.
- Paucar-Santivañez, A. P. (2020). Propiedades psicométricas de la escala de factores psicosociales en el trabajo en trabajadores peruanos del sector minero. (Licenciatura en Psicología). Universidad San Ignacio de Loyola (USIL), Lima Perú. Disponible en: http://repositorio.usil.edu.pe/bitstream/USIL/11233/3/2020_Paucar%20Santiva%C3%B1ez.pdf
- Pearson, R., Mundfrom, D., & Piccone, A. (2013). A comparison of ten methods for determining the number of factors in exploratory factor analysis. *Multiple Linear Regression Viewpoints*, 39(1), 1-15.
- Penfield, R. D., & Giacobbi, P. R., Jr. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8(4), 213-225. https://doi.org/10.1207/s15327841mpee0804_3
- Pines, A., & Aronson, E. (1988). *Career burnout: causes and cure*. Nueva York: The Free Press.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29, 489-497. <https://doi.org/10.1002/nur.20147>

- Polit, D. F., Beck, C. T., & Owen, S. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*, 459-467.
<https://doi.org/10.1002/nur.20199>
- Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills, 105*(3 Pt 1), 997–1014.
<https://doi.org/10.2466/pms.105.3.997-1014>
- Poortinga, Y. H. (1989). Equivalence of cross-cultural data: an overview of basic issues. *International Journal of Psychology, 24*(6), 737-756.
- Pope, C., Ziebland, S., & Mays, N. (2000). Qualitative research in health care. Analysing qualitative data. *BMJ Clinical Research Education, 320*(7227), 114–116.
<https://doi.org/10.1136/bmj.320.7227.114>
- Porta, M. (2014). *A dictionary of epidemiology* (6th. Edition). New York: Oxford University Press
- Potocka A. (2012). Narzędzia kwestionariuszowe stosowane w diagnozie psychospołecznych zagrożeń zawodowych: Przegląd metod [Questionnaires for assessment of work-related psychosocial hazards: A review of diagnostic tools]. *Medycyna Pracy, 63*(2), 237–520.
- Preckel, F., & Brunner, M. (2017). Nomological Nets. En: V. Zeigler-Hill, & T. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–4). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-28099-8_1334-1
- Price, L. R. (2017). *Psychometric methods: theory and practice*. New York: Guilford Press.
- Prince, M. (2008). Measurement validity in cross-cultural comparative research. *Epidemiology and Psychiatric Sciences, 17*(3), 211-220.

- Proper, K. I., van den Heuvel S. G., De Vroome, E. M., Hildebrandt, V. H., & van der Beek A. J. (2006). Dose-response relation between physical activity and sick leave *British Journal of Sports Medicine*, *40*, 173-178. <https://doi.org/10.1136/bjsm.2005.022327>
- Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior*, *77*, 168-185. <https://doi.org/10.1016/j.jvb.2004.04.006>
- Quick, J.C. (1999a). Occupational Health Psychology: historical roots and future directions. *Health Psychology*, *18*, 82-88.
- Quick, J.C. (1999b). Occupational Health Psychology: The convergence of Health and Clinical Psychology with Public Health and Preventive Medicine in an organizational context. *Professional Psychology: Research and Practice*, *30*, 123-128.
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Raiche, G., & Magis, D. (2020). nFactors: parallel analysis and other non graphical solutions to the Cattell Scree Test. R package version 2.4.1. <https://CRAN.R-project.org/package=nFactors>
- Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517-529. <https://doi.org/10.1037//0021-9010.87.3.517>
- Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized Conditional SEM: A case for conditional reliability. *Applied Psychological Measurement*, *31*(3), 169–180. <https://doi.org/10.1177/0146621606291569>
- Ramírez, A. (2012). Servicios de salud ocupacional. *Anales de la Facultad de Medicina*, *73*(1), 63-69.
- Reina-Tamayo, A. M., Bakker, A. B., & Derks, D. (2017). Episodic demands, resources, and engagement. *Journal of Personnel Psychology*, *16*(3), 125-136. <https://doi.org/10.1027/1866-5888/a000177>

- Ren, C., Li, X., Yao, X., Pi, Z., & Qi, S. (2019). Psychometric properties of the Effort-Reward Imbalance Questionnaire for Teachers (Teacher ERIQ). *Frontiers in psychology, 10*, 2047. <https://doi.org/10.3389/fpsyg.2019.02047>
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., & Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science, 334*(6062), 1518–1524. <https://doi.org/10.1126/science.1205438>
- Reshef, D. N., Reshef, Y. A., Sabeti, P. C., & Mitzenmacher, M. (2018). An empirical study of the maximal and total information coefficients and leading measures of dependence. *Annals of Applied Statistics, 12*(1), 123-155. <https://doi.org/10.1214/17-AOAS1093>
- Revelle, W. (2020) psych: Procedures for personality and psychological research. R package version 2.0.12. <https://CRAN.R-project.org/package=psych>
- Revelle, W. (2021). psych: Procedures for Personality and Psychological Research. R package version 2.1.6. <https://CRAN.R-project.org/package=psych>
- Rick, J. & Briner, R. B. (2000). Psychosocial risk assessment: problems and prospects. *Occupational Medicine, 50*, 310-314. <https://doi.org/10.1093/occmed/50.5.310>
- Rigotti, T., Schyns, B., & Mohr, G. (2008). A short version of the occupational self-efficacy scale: Structural and construct validity across five countries. *Journal of Career Assessment, 16*(2), 238-255. <https://doi.org/10.1177/F1069072707305763>
- Ríos, I. (2017). Un acercamiento a la legibilidad de textos relacionados con el campo de la salud. Chasqui. *Revista Latinoamericana de Comunicación, 135*, 253-273
- Robitzsch, A. (2019). Supplementary Item Response Theory Models Version. R-project, Package 'sirt' manual. <https://cran.r-project.org/web/packages/sirt/sirt.pdf>

- Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education, 5*, 177. <https://doi.org/10.3389/feduc.2020.589965>
- Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods, 25*(1), 6–14. <https://doi.org/10.1177/1094428120968614>
- Rosário, S., Fonseca, J. A., Nienhaus, A., & Torres, J. (2016). Standardized assessment of psychosocial factors and their influence on medically confirmed health outcomes in workers: a systematic review. *Journal of Occupational Medicine and Toxicology, 11*(19). Disponible en: <https://occup-med.biomedcentral.com/articles/10.1186/s12995-016-0106-9>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Rousselet, G.A., Pernet, C.R., & Wilcox, R.R. (2019). A practical introduction to the bootstrap: a versatile method to make inferences by using data-driven simulations (Preprint). <https://psyarxiv.com/h8ft7/>
- Roussos, L. A., & Ozbek, O. Y. (2006). Formulation of the DETECT population parameter and evaluation of DETECT estimator bias. *Journal of Educational Measurement, 43*, 215–243.
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research, 27*(2), 94–104. <https://doi.org/10.1093/swr/27.2.94>
- Ruedin, D. (2020). agrmt: Calculate concentration and dispersion in ordered rating scales. R package version 1.42.4. <https://CRAN.R-project.org/package=agrmt>
- Rupp, A. A. (2013). A systematic review of the methodology for person fit research in Item Response Theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling, 55*(1), 3-38.

- Rutkowski, L., Svetina, D., & Liaw, Y.L. (2019). Collapsing Categorical Variables and Measurement Invariance, *Structural Equation Modeling*, 26(5), 790-802.
<https://doi.org/10.1080/10705511.2018.1547640>
- Salminen, S., Kouvonen, A., Koskinen, A., Joensuu, M., & Väänänen, A. (2014). Is a single item stress measure independently associated with subsequent severe injury: a prospective cohort study of 16,385 forest industry employees. *BMC Public Health*, 14, 543. <https://doi.org/10.1186/1471-2458-14-543>
- Samejima, F. (1980). Research on the multiple-choice test item in Japan toward the validation of mathematical models. Universidad de Michigan, Psychological Sciences Division, Office of Naval Research, Disponible en: <https://files.eric.ed.gov/fulltext/ED193313.pdf>
- Sanchez J.I., Spector P.E., Cooper C.L. (2006). Frequently ignored methodological issues in cross-cultural stress research. In: Wong P.T.P., Wong L.C.J. (eds), *Handbook of Multicultural Perspectives on Stress and Coping. International and Cultural Psychology* (pp. 187-201). Springer, Boston, MA. https://doi.org/10.1007/0-387-26238-5_9
- Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, 16, 561-582.
<https://doi.org/10.1080/10705510903203433>
- Sarstedt, M., Bengart, P., Shaltoni, A. M., & Lehmann, S. (2018). The use of sampling methods in advertising research: a gap between theory and practice. *International Journal of Advertising*, 37(4), 650-663. <https://doi.org/10.1080/02650487.2017.1348329>
- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: a comparison of estimators. *Structural Equation Modeling*, 21(2), 167-180. <https://doi.org/10.1080/10705511.2014.882658>

- Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: a comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(2), 167-180.
<https://doi.org/10.1080/10705511.2014.882658>
- Sato, T. & Morimoto, U. (1976). Sentaku-shi keishiki tesuto kaitou bunpu no bunseki [Analyzing endorsement distribution of selected-response items]. In Proceedings of the 4th Annual Meeting of the Behaviometric Society of Japan, Behaviometric Society of Japan. Tokio.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.
- Savalei, V. (2014). Understanding robust corrections in structural equation modeling. *Structural Equation Modeling*, 21(1), 149–160. <https://doi.org/10.1080/10705511.2013.824793>
- Savalei, V., & Rhemtulla, M. (2013). The performance of robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, 66(2), 201–223.
<https://doi.org/10.1111/j.2044-8317.2012.02049.x>
- Sax, L.J., Gilmartin, S.K., & Bryant, A.N. (2003). Assessing response rates and nonresponse bias in web and paper surveys. *Research in Higher Education*, 44, 409-432.
<https://doi.org/10.1023/A:1024232915870>
- Schabracq, M. J. (2003). Everyday well-being and stress in work and organisations. En M. J. Schabracq, J. A. M. Winnubst & C. L. Cooper (Eds.), *The Handbook of Work and Health Psychology*, 2nd (pp. 9-36). Chichester: John Wiley & Sons.
<https://doi.org/10.1002/0470013400.ch2>
- Schaufeli, W. (2004). The future of occupational health psychology. *Applied Psychology: An International Review*, 53(4), 502-517. <https://doi.org/10.1111/j.1464-0597.2004.00184.x>

- Schaufeli, W. B., & Taris, T. W. (2014). A critical review of the job demands-resources model: Implications for improving work and health. In G. F. Bauer & O. Hämmig (Eds.), *Bridging occupational, organizational and public health: A transdisciplinary approach* (pp. 43-68). Dordrecht: Springer Science + Business Media. https://doi.org/10.1007/978-94-007-5640-3_4
- Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. (2002). The measurement of engagement and burnout: a two-sample confirmatory factor analytic-approach. *Journal of Happiness Studies*, 3(1), 71-92.
- Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (2019). An ultra-short measure for work engagement: The UWES-3 validation across five countries. *European Journal of Psychological Assessment*, 35(4), 577–591. <https://doi.org/10.1027/1015-5759/a000430>
- Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & De Witte, H. (2019). An ultra-short measure for work engagement: The UWES-3 validation across five countries. *European Journal of Psychological Assessment*, 35(4), 577–591. <https://doi.org/10.1027/1015-5759/a000430>
- Schaufeli, W.B. (2013). What is engagement? In C. Truss, K. Alfes, R. Delbridge, A. Shantz, & E. Soane (Eds.), *Employee Engagement in Theory and Practice* (pp. 1-37). London: Routledge.
- Schmoldt, R. A, Freeborn, D. K., & Klevit, H. D. (1994). Physician burnout: recommendations for HMO managers. *HMO Practice*, 8(2), 58–63.
- Scholtz, S. E., de Klerk, W, & de Beer, L. T. (2020). The use of research methods in psychological research: A systematised review. *Frontiers in Research Metrics and Analytics*, 5, 1. <https://doi.org/10.3389/frma.2020.00001>

- Schroeders, U., Schmidt, C., & Gnambs, T. (2022). Detecting careless responding in survey data using Stochastic Gradient Boosting. *Educational and Psychological Measurement*, 82(1), 29-56. <https://doi.org/10.1177/00131644211004708>
- Schutte, N.S., Malouff, J.M., Hall, L.E., Haggerty, D.J., Cooper, J.T., Golden, C.J., et al. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences*, 25, 167-177. [https://doi.org/10.1016/s0191-8869\(98\)00001-4](https://doi.org/10.1016/s0191-8869(98)00001-4)
- Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31(12), 1428–1441. <https://doi.org/10.1037/pas0000623>
- Setterlind, S., & Larsson, G. (2015). The stress profile: A psychosocial approach to measuring stress. *Stress Medicine*, 11, 85-92. <https://doi.org/10.1002/smi.2460110116>
- Sheatsley, P. B. (1983). Questionnaire construction and item writing. In P. H. Rossi, J.D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 195–230). San Diego: Academic Publishers.
- Shoji, K., Cieslak, R., Smoktunowicz, E., Rogala, A., Benight, C. C., & Luszczynska, A. (2016). Associations between job burnout and self-efficacy: a meta-analysis. *Anxiety, Stress & Coping*, 29(4), 367-386. <https://doi.org/10.1080/10615806.2015.1058369>
- Siegrist, J. & Peter, R. (2000). The Effort- Reward Imbalance Model. *Occupational Medicine*. 1, 83-86.
- Siegrist, J. & Place, J. (2000). Social exchange and health: proposed sociological framework. *Social Science Medicine*, 51, 1283-1293. <https://doi.org/10.1016/b978-0-12-800951-2.00009-1>
- Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1, 27–41. <https://doi.org/10.1037/1076-8998.1.1.27>

- Siegrist, J., Wege, N., Pühlhofer, F., & Wahrendorf, M. (2009). A short generic measure of work stress in the era of globalization: effort-reward imbalance. *International Archives of Occupational and Environmental Health*, 82(8), 1005–1013. <https://doi.org/10.1007/s00420-008-0384-3>
- Signorell, A. et mult. al. (2021). DescTools: Tools for descriptive statistics. R package version 0.99.43.
- Sijtsma, K., & Hemker, B. T. (2000). A Taxonomy of IRT models for ordering persons and items using simple sum scores. *Journal of Educational and Behavioral Statistics*, 25(4), 391–415. <https://doi.org/10.3102/10769986025004391>
- Sijtsma, K., & Van der Ark, L. A. (2003). Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research*, 38, 505–528. https://doi.org/10.1207/s15327906mbr3804_4
- Sijtsma, K., & van der Ark, L. A. (2017). A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. *The British Journal of Mathematical and Statistical Psychology*, 70(1), 137–158. <https://doi.org/10.1111/bmsp.12078>
- Sijtsma, K., Meijer, R. R., & Andries van der Ark, L. (2011). Mokken scale analysis as time goes by: An update for scaling practitioners. *Personality and Individual Differences*, 50(1), 31–37. <https://doi.org/10.1016/j.paid.2010.08.016>.
- Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the unidimensionality of psychological scales: Using multiple criteria from factor analysis. *Social Indicator Research*, 102, 443–461. <https://doi.org/10.1007/s11205-010-9682-8>
- Smith, E. V., Jr., Wakely, M. B., De Kruif, R. E. L., & Swartz, C. W. (2003). Optimizing rating scales for self-efficacy (and other) research. *Educational and Psychological Measurement*, 63(3), 369–391. <https://doi.org/10.1177/0013164403063003002>

- Smits, I. A. M., Timmerman, M. E., & Meijer, R. R. (2012). Exploratory Mokken scale analysis as a dimensionality assessment tool: why scalability does not imply unidimensionality. *Applied Psychological Measurement, 36*(6), 516–539. <https://doi.org/10.1177/0146621612451050>
- Solano-Flores, Guillermo & Milbourn, Tamara (2016). Capacidad evaluativa, validez cultural y validez consecuencial en PISA. *Revista Electrónica de Investigación y Evaluación Educativa, 22*, 1-17. <http://dx.doi.org/10.7203/relieve.22.1.8281>
- Solíz, M. (2015). The dilemma of combining positive and negative items in scales. *Psicothema, 27*(2), 192-199. <https://dx.doi.org/0.7334/psicothema2014.266>
- Soper, D.S. (2017). A-priori sample size calculator for structural equation models [Software]. Disponible en <http://www.danielsoper.com/statcalc>.
- Sörbom, D. (1989). Model modification. *Psychometrika, 54*, 371–384. <https://doi.org/10.1007/bf02294623>
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology, 15*, 72–101.
- Speed, T. (2011). A correlation for the 21st Century. *Science, 334*(6062), 1502-1503. <https://doi.org/10.1126/science.1215894>
- Spielberger, C., & Reheiser, E. (2009). Assesment of emotions: Anxiety, anger, depression and curiosity. *International Association of Applied Psychology, 1* (3), 271-302. <https://doi.org/10.1111/j.1758-0854.2009.01017.x>
- Stansfeld, S., & Candy, B. (2006). Psychosocial work environment and mental health--a meta-analytic review. *Scandinavian Journal of Work, Environment & Health, 32*(6), 443–462. <https://doi.org/10.5271/sjweh.1050>

- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: toward a unified strategy. *The Journal of applied psychology, 91*(6), 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Steenkamp, J. B. E. M., & Baumgartner, J. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78-90. <https://doi.org/10.1086/209528>
- Stenfors, C. U., Magnusson Hanson, L., Oxenstierna, G., Theorell, T., & Nilsson, L. G. (2013). Psychosocial working conditions and cognitive complaints among Swedish employees. *PloS one, 8*(4), e60637. <https://doi.org/10.1371/journal.pone.0060637>
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*, 589–617.
- Stuckey, H. L. (2015). The second step in data analysis: Coding qualitative research data. *Journal of Social Health and Diabetes, 3*(1), 7-10. <https://doi.org/10.4103/2321-0656.140875>
- Suárez, A. (2013). Adaptación de la Escala de estrés laboral de la OIT-OMS en trabajadores de 25 a 35 años de edad de un Contact Center de Lima. *Revista PsiqueMag*. Disponible en: <http://ojs.ucvlima.edu.pe/index.php/psiquemag/article/download/8/8>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295-312. [http://dx.doi.org/10.1016/0959-4752\(94\)90003-5](http://dx.doi.org/10.1016/0959-4752(94)90003-5)
- Szigriszt Pazos, F. (1993). *Sistemas predictivos de legibilidad del mensaje escrito: fórmula de perspicuidad*. Tesis doctoral, Universidad Complutense de Madrid. Facultad de Ciencias de la Comunicación.
- Tabanelli, M. C., Depolo, M., Cooke, R. M. T., Sarchielli, G., Bonfiglioli, Mattioli, S., Violante, F. S. (2008). Available instruments for measurement of psychosocial factors in the work

environment. *International Archives of Occupational Environment Health*, 82, 1–12.

<https://doi.org/10.1007/s00420-008-0312-6>

Tastle, W., & Wierman, M. (2007). Consensus and dissent: A measure of ordinal dispersion.

International Journal of Approximate Reasoning, 45(3), 531-545.

te Nijenhuis, J., & van der Flier, F. J. R. (2009). Bias research in the Netherlands: Review and implications. *European Journal of Psychological Assessment*, 15, 165-175.

<https://doi.org/10.1027//1015-5759.15.2.165>

Teresi, J. A. (2006). Overview of quantitative measurement methods: equivalence, invariance, and differential item functioning in health applications. *Medicine Care*, 44(3), 39–49.

Teresi, J. A., Ramirez, M., Lai, J.-S., & Silver, S. (2008). Occurrences and sources of Differential Item Functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science Quarterly*, 50(4), 538. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2822361/>

Theorell, T., Hammarström, A., Aronsson, G., Träskman Bendz, L., Grape, T., Hogstedt, C., Marteinsdottir, I., Skoog, I., & Hall, C. (2015). A systematic review including meta-analysis of work environment and depressive symptoms. *BMC Public Health*, 15, 738.

<https://doi.org/10.1186/s12889-015-1954-4>

Toderi, S., & Sarchielli, G. (2016). Psychometric Properties of a 36-Item Version of the "Stress Management Competency Indicator Tool". *International Journal of Environmental Research and Public Health*, 13(11), 1086. <https://doi.org/10.3390/ijerph13111086>

Tong, X., & Bentler, P. M. (2013). Evaluation of a new mean scaled and moment adjusted test statistic for SEM. *Structural Equations Modeling*, 20(1), 148–156.

<https://doi.org/10.1080/10705511.2013.742403>

- Torchiano M (2020). *effsize: Efficient Effect Size Computation*. R package version 0.8.0. Disponible en: <https://CRAN.R-project.org/package=effsize>
- Torres-Frías, J., Moreno-Bayardo, M., & Jiménez-Mora, J. (2018). Aportes de lectores y lectoras de tesis doctoral como mediación pedagógica en la formación de personas investigadoras. *Revista Educación* 42(1), 194-214. <https://doi.org/10.15517/revedu.v42i1.22970>
- Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (No. PRS-984). Princeton, N J, USA: Educational Testing Service.
- Unda, S., Uribe, F., Jurado, S., García, M., Tovalín, H., & Juárez, A. (2016). Elaboración de una escala para valorar los factores de riesgo psicosocial en el trabajo de profesores universitarios. *Revista de Psicología del Trabajo y de las Organizaciones*, 32(2), 67-74. <https://dx.doi.org/10.1016/j.rpto.2016.04.004>
- Universidad Peruana Cayetano Heredia (s/f). Riesgos Psicosociales en el Trabajo y Psicología de la Salud Ocupacional. Consultado: 15/05/2019. Disponible en: <http://www.upch.edu.pe/famed/posgrado/es/diplomados/riesgos-psicosociales-en-el-trabajo-y-psicologia-de-la-salud-ocupacional>
- Urbina, S. (2014). *Essentials of psychological testing (2d. ed.)*. Hoboken, NJ: Wiley & Sons.
- Uribe Prado, J., Gutiérrez Amador, J., & Amézquita Pino, J. (2020). Critique on the psychometric properties of a measurement scale for psychosocial risk factors proposed in NOM 035 of the STPS in Mexico. *Contaduría y Administración*, 65(1), e147. <https://doi.org/10.22201/fca.24488410e.2019.1569>
- Urrutia, M., Barrios, S., Gutiérrez, M., & Mayorga, M. (2014). Métodos óptimos para determinar la validez de contenido. *Educación Médica Superior*, 38(3). Disponible en: <http://www.ems.sld.cu/index.php/ems/rt/printerFriendly/301/192>

- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58, 6-20.
<https://doi.org/10.1177/0013164498058001002>
- Vallejo-Medina, P., Gómez-Lugo, M., Marchal-Bertrand, L., Saavedra-Roa, A., Soler, F., & Morales, A. (2017). Developing guidelines for adapting questionnaires into the same language in another culture. *Terapia Psicológica*, 35(2), 159-172.
- van Abswoude, A. A. H., van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT models. *Applied Psychological Measurement*, 28(1), 3–24. <https://doi.org/10.1177/0146621603259277>
- van der Ark, L. A. (2005). Stochastic ordering of the latent trait by the sum score under various polytomous IRT models. *Psychometrika*, 70, 283–304. <https://doi.org/10.1007/s11336-000-0862-3>
- Van der Ark, L. A. (2012). New developments in Mokken scale analysis in R. *Journal of Statistical Software*, 48. <https://www.jstatsoft.org/article/view/v048i05>
- van der Doef, M., & Maes, S. (1999). The Job Demand-Control (-Support) model and psychological well-being: A review of 20 years of empirical research. *Work & Stress*, 13(2), 87-114.
<https://doi.org/10.1080/026783799296084>
- Van Dijk, T. K., Datema, F., Piggen, A-L. J. H. F., Welten, S. C. M., & Van de Vijver, F. J. R. (2009) Acquiescence and extremity in cross-national surveys: domain dependence and country-level correlates. In A. Gari, & K. Mylonas (Ed.s) *Quod erat Demonstrandum: From Herodotus' Ethnographic Journeys to Cross-Cultural Research* (pp. 149–158). Athens: Pedio Books Publishing.

- van Gelder, M. M. H. J., Bretveld, R. W., & Roeleveld, N. (2010). Web-based Questionnaires: The future in epidemiology? *American Journal of Epidemiology*, *172*(11), 1292–1298.
<https://doi.org/10.1093/aje/kwq291>
- Van Ginkel, J. R., & Van der Ark, L. A. (2005). SPSS syntax for missing value imputation in test and questionnaire data. *Applied Psychological Measurement*, *29*, 152–153.
<https://doi.org/10.1177/0146621603260688>
- Van Ginkel, J. R., Van der Ark, L. A. & Sijtsma, K. (2007). Multiple imputation of test and questionnaire data and influence on psychometric results. *Multivariate Behavioral Research*, *42*, 387–414. <https://doi.org/10.1080/00273170701360803>
- Van Raaij, W. F. (1978). Cross-cultural research methodology as a case of construct validity. En H. K. Hunt (Ed.), *Advances in Consumer Research* (pp. 693-701). Ann Arbor: Association for Consumer Research.
- van Vegchel, N., de Jonge, J., Bosma, H. and Schaufeli, W. (2005). Reviewing the effort–reward imbalance model: drawing up the balance of 45 empirical studies. *Social Science & Medicine*, *60*(5), 1117–1131. <https://doi.org/10.1016/j.socscimed.2004.06.043>
- Van Wijk, C.H. (2021). Usefulness of the English version of the Stress Overload Scale in a sample of employed South Africans. *African Journal of Psychological Assessment*, *3*(0), a41.
<https://doi.org/10.4102/ajopa.v3i0.41>
- Varela, M., & Vives, T. (2016). Autenticidad y calidad en la investigación educativa cualitativa: multivocalidad. *Investigación en Educación Médica*, *5*(19), 191-198.
<https://doi.org/10.1016/j.riem.2016.04.006>
- Voorhees, C. M., Brady, M. K., Calantone, R., & Ramirez, E. (2016). Discriminant validity testing in marketing: an analysis, causes for concern, and proposed remedies. *Journal of the Academy of Marketing Science*, *44*(1), 119–134. <https://doi.org/10.1007/s11747-015-0455-4>

- Wang, L.-W., Miller, M. J. Schmitt, M. R., & Wen, F. K. (2013). Assessing readability formula differences with written health information materials: Application, results, and recommendations. *Research in Social and Administrative Pharmacy, 9*, 503–516.
<http://dx.doi.org/10.1016/j.sapharm.2012.05.009>
- Wang, Y., Li, Y., Cao, H., Xiong, M., Shugart, Y. Y., & Jin, L. (2015). Efficient test for nonlinear dependence of two continuous variables. *BMC Bioinformatics, 16*, 260.
<https://doi.org/10.1186/s12859-015-0697-7>
- Wanous, J. P., Reichers, A. E., & Hudy, M. J. (1997). Overall job satisfaction: how good are single-item measures? *Journal of Applied Psychology, 82*(2), 247-252. <https://doi.org/10.1037//0021-9010.82.2.247>
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology, 44*(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Watson, R., Egberink, I. J. L., Kirke, L., Tendeiro, J. N. & Doyle, F. (2018) What are the minimal sample size requirements for Mokken scaling? An empirical example with the Warwick-Edinburgh Mental Well-Being Scale. *Health Psychology and Behavioral Medicine, 6*(1), 203-213. <https://doi.org/10.1080/21642850.2018.1505520>
- Wengraf, T. (2001). *Qualitative Research Interviewing*. London: SAGE.
- Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84*(3), 608-618. <http://dx.doi.org/10.1037/0022-3514.84.3.608>
- Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*(6), 476-487.
<https://doi.org/10.1016/j.elerap.2010.07.003>

- Whitmore, M. L., & Schumacker, R. E. (1999). A comparison of logistic regression and analysis of variance differential item functioning detection methods. *Educational and Psychological Measurement, 59*, 910–927. <https://doi.org/10.1177/00131649921970251>
- Widerszal-Bazyl, M., & Cieslak, R. (2000). Monitoring psychosocial stress at work: development of the Psychosocial Working Conditions Questionnaire. *International Journal of Occupational Safety and Ergonomics, 1*, 59-70. <https://doi.org/10.1080/10803548.2000.11105108>
- Wilson, A., Wissing, M.P., & Schutte, L. (2018). Validation of the stress overload scale and stress overload scale – Short form among a Setswana-speaking community in South Africa. *South African Journal of Psychology, 48*(1), 21–31. <https://doi.org/10.1177/0081246317705241>
- Winton, B., G., & Sabol, M. A. (2021). A multi-group analysis of convenience samples: free, cheap, friendly, and fancy sources. *International Journal of Social Research Methodology*, <https://doi.org/10.1080/13645579.2021.1961187>
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934. <http://doi.org/10.1177/0013164413495237> .
- Wong, C. S. & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *Leadership Quarterly, 13*(3), 243-274. [https://doi.org/10.1016/s1048-9843\(02\)00099-1](https://doi.org/10.1016/s1048-9843(02)00099-1)
- Wood, S., Ghezzi, V., Barbaranelli, C., Di Tecco, C., Fida, R., Farnese, M. L., Ronchetti, M., & Iavicoli, S. (2019). Assessing the risk of stress in organizations: Getting the measure of organizational-level stressors. *Frontiers in Psychology, 10*, 2776. <https://doi.org/10.3389/fpsyg.2019.02776>

- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186-191.
<https://doi.org/10.1007/s10862-005-9004-7>
- World Health Organization (2007). *Raising Awareness of Stress at Work in Developing Countries*. (Protecting Workers' Health Series No. 6). Geneva: World Health Organization
- World Health Organization (s/f). Stress in the workplace. Disponible en:
https://www.who.int/occupational_health/topics/stressatwp/en/
- Ximénez, C. (2006). A Monte Carlo study of recovery of weak factor loadings in confirmatory factor analysis. *Structural Equation Modeling*, 13(4), 587–614.
https://doi.org/10.1207/s15328007sem1304_5
- Ximénez, M. C., & García, A. G. (2005). Comparación de los métodos de estimación de máxima verosimilitud y mínimos cuadrados no ponderados en el análisis factorial confirmatorio mediante simulación Monte Carlo. *Psicothema*, 17(3), 528-535.
- Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling*, 17, 392–423.
- Yoshida, K., & Bartel, A. (2021). tableone: Create 'Table 1' to describe baseline characteristics with or without propensity score weights. R package versión 0.13.0. <https://CRAN.R-project.org/package=tableone>
- Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A. R. (2015). Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences*, 4(2), 165–178.
<https://doi.org/10.15171/jcs.2015.017>
- Abbas, R. A., Zalat, M. M., & Ghareeb, N. S. (2013). Non-fatal occupational injuries and safety climate: A cross-sectional study of construction building workers in Mit-Ghamr City, Dakahlia

Governorate. *Egypt Open Journal of Safety Science and Technology*, 3(4), 69–79.

<https://doi.org/10.4236/ojsst.2013.34009>

Zhang, J. (2007). Conditional covariance theory and DETECT for polytomous items. *Psychometrika*, 72(1), 69-91. <https://doi.org/10.1007/s11336-004-1257-7>

Zhang, J., & Stout, W. (1999b). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64(2), 213-249.

<https://doi.org/10.1007/BF02294536>

Zhang, J., & Stout, W. F. (1999a). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64, 129–152. <https://doi.org/10.1007/BF02294532>

Żołnierczyk-Zreda, D., & Holas, P. (2018). Psychospołeczne warunki pracy a depresja i zaburzenia depresyjne: przegląd badań [Psychosocial working conditions and major depression or depressive disorders: Review of studies]. *Medycyna pracy*, 69(5), 573–581.

<https://doi.org/10.13075/mp.5893.00703>

Apéndice

Apéndice 1

The Self-Assessed Literacy Index (SALI; Olson et al., 2011)

Considerando el idioma castellano que tú manejas, ¿qué tan bueno eres..?

... para entenderlo cuando te hablan?	Bajo	No tan bien	Bien	Muy bien
... para hablarlo (expresarte oralmente)?	Bajo	No tan bien	Bien	Muy bien
... para leerlo (comprender lo que lees)?	Bajo	No tan bien	Bien	Muy bien
... para escribirlo (expresarte por escrito)?	Bajo	No tan bien	Bien	Muy bien

¿Con qué frecuencia lees...?

... Periódicos o revistas?	Muy poco	Algunas veces	Frecuente	A diario
... Libros?	Muy poco	Algunas veces	Frecuente	A diario
... Otros documentos (notas, reportes, etc.)?	Muy poco	Algunas veces	Frecuente	A diario

¿Con qué frecuencia tiene problemas para comprender información o instrucciones escritas?	Siempre	Frecuente	A veces	Raramente	Nunca
---	----------------	------------------	----------------	------------------	--------------

Contenido del formulario de Consentimiento Informado en la Plataforma web

Factores Psicosociales en el Trabajo

CONSENTIMIENTO PARA PARTICIPAR EN EL PROYECTO

PROPÓSITO DEL ESTUDIO:

Evaluar el funcionamiento de un cuestionario para identificar aspectos psicosociales relacionados con el estrés y el bienestar laboral.

PROCEDIMIENTO:

El investigador del presente proyecto le proporcionará las preguntas del material mediante un cuestionario, y su participación consistirá en responder a todas las preguntas de manera honesta.

RIESGOS POTENCIALES Y MOLESTIAS:

Anticipamos un mínimo riesgo al participar en esta investigación. Con respecto a los cuestionarios, posiblemente usted encuentre que algunas preguntas le produzcan alguna incomodidad para responderlas en ciertos temas. En este caso le recordamos que:

- a. toda información que proporcione será manejada bajo estricta confidencialidad, y por lo tanto:
- b. le solicitamos no anotar su nombre en ninguna parte del cuestionario. Los archivos de información estarán bajo acceso restringido. La información se tratará únicamente con fines de análisis estadísticos y sin alguna manera de identificarlo a usted.

BENEFICIOS POTENCIALES PARA LOS SUJETOS Y PARA LA SOCIEDAD:

No se ofrecerán beneficios económicos u otro tipo directo por su participación. Sin embargo, usted puede adquirir conocimientos organizados sobre su institución laboral, y en caso de solicitarlo, se le entregará un informe confidencial e individualizado de sus resultados personales. Para ello deberá solicitarlo a los investigadores (abajo, mediante el correo de contacto) indicando un email. No hay ningún otro beneficio adicional específico para usted. Esta investigación intenta incrementar el conocimiento del tema de los factores psicosociales y el estrés y bienestar laboral.

CONFIDENCIALIDAD DE LOS DATOS

La información sobre el número de folio y su nombre y/o firma en este formulario de consentimiento informado, se mantendrán confidencialmente, y no aparecerá en algún medio esta información que pudiera identificarlo.

PARTICIPACION O RETIRO:

Usted puede decidir si participa o no en este estudio sin ninguna consecuencia de ningún tipo. En este sentido, tiene completa libertad de dejar de participar en cualquier momento. También, el investigador lo puede retirar de esta investigación si las circunstancias lo justifican.

IDENTIFICACION DE LOS INVESTIGADORES:

Si tiene cualquier pregunta respecto a esta investigación, por favor siéntase libre de contactar a: Mtro. César Merino Soto UAEM, correo: cesar.merinosto@uaem.edu.mx

DERECHOS DE LOS PARTICIPANTES EN INVESTIGACION: Usted puede retirar su consentimiento en cualquier momento y detener su participación sin alguna penalidad o consecuencia personal. Usted no está renunciando a cualquier demanda, derecho, recurso legal por su participación en este estudio.

***Obligatorio**

1. Entiendo los procedimientos descritos anteriormente. Mis preguntas han sido contestadas a mi satisfacción y estoy de acuerdo en participar en este estudio *

Marca solo un óvalo.

Sí

No *Ir a la sección 9 (Final)*

2. Si quisiera un informe de la encuesta, informar de un email aquí, o en caso contrario, deje en blanco:

Estructura de la Hoja Demográfica

Instrucciones:

Muchas gracias por el tiempo que va a dedicar a responder a este cuestionario, le recordamos que es confidencial y la falta de llenado en alguna pregunta lo invalida totalmente, por lo que agradecemos su cooperación y sinceridad en TODAS las preguntas.

En esta parte del cuestionario debe reflejar algunos datos personales. Su objetivo es poder agrupar sus respuestas con la de otros profesionales de características similares a las suyas para fines de análisis estadístico. Según sean sus datos en cada pregunta conteste seleccionando sobre los números del lado derecho.

Estructura

Tipo de información	Formato de pregunta	Contenido	Formato de respuesta
Sexo	Nominal dicotómica	<ul style="list-style-type: none"> • Hombre • Mujer 	Única
Fecha de nacimiento	Fecha	-	Única
Estado civil	Multinomial	<ul style="list-style-type: none"> • Soltero • Casado • unión libre (conviviente) • divorciado • viudo 	Única
Convivencia	Multinomial	<ul style="list-style-type: none"> • Solo • solo con mascota • con pareja o familiares • con amigos • otros 	Múltiple
Adscripción religiosa	Multinomial	<ul style="list-style-type: none"> • Católica • Cristiana • Testigo de Jehová • Judía • Mormona • Ateo • Otros 	Única
Número de hijos	Abierta	Espacio libre	Libre
Personas económicamente dependientes en el hogar	Abierta	Espacio libre	
¿Cuántas personas dependen económicamente de usted?	Abierta	Espacio libre	Libre
Último grado de instrucción alcanzado	Multinomial	<ul style="list-style-type: none"> • Primaria (no terminada o terminada) • secundaria terminada • bachillerato/carrera técnica • licenciatura (universidad terminada) • postgrado (terminado) • otros 	Única

Tipo de información	Formato de pregunta	Contenido	Formato de respuesta
Habitación	Multinomial	<ul style="list-style-type: none"> • Casa propia • casa con renta (alquilada) • prestada (de familiares o amigos) • otros. 	Única
Servicios con los que se cuenta	Multinomial	<ul style="list-style-type: none"> • Servicios de drenaje • Agua • Luz/electricidad • Teléfono fijo • Servicio de TV de Paga • Aire acondicionado • Ninguno 	Múltiple
Seguridad del vecindario	Ordinal	<ul style="list-style-type: none"> • Totalmente seguro • Más o menos seguro/inseguro • Totalmente Inseguro 	Única

Sobre las características de su trabajo

Tipo de información	Formato de pregunta	Contenido	Formato de respuesta
Puesto actual	Abierta		Libre
Jerarquía dentro de la empresa	Multinomial	<ul style="list-style-type: none"> • Director • Gerente • supervisor o jefe de área • Administrativo sin personal a cargo (trabajo no manual) • Ventas sin personal a cargo (trabajo no manual) • Empleado operativo (trabajo manual) • Otros 	Única
Tiempo trabajado en el puesto actual	Abierta	Espacio libre	Libre
Horas promedio de trabajo en el puesto actual	Abierta	Espacio libre	Libre
Tipo de contrato	Multinomial	<ul style="list-style-type: none"> • Planta/definitivo. • Eventual 	Única
Tiempo de trabajo en la empresa	Abierta	- Espacio libre	Libre
Rango de salario mensual	Ordinal	<ul style="list-style-type: none"> • No recibo un salario • Hasta 2,000 pesos • De 2,001 a 4,000 pesos • De 4,001 a 8,000 pesos • De 8,001 a 14,000 pesos • De 14,001 en adelante 	Única

Eventos personales

Tipo de información	Formato de pregunta	Contenido	Formato de respuesta
Promedio de horas para dormir	Ordinal	<ul style="list-style-type: none"> • Menos de 6 hrs. • 6 a 7 hrs. • 7 a 8 hrs. • Más de 8 hrs. 	Única
Medio de transporte para ir a trabajar	Multinomial	<ul style="list-style-type: none"> • Auto propio • En el auto de un conocido • Compañeros • Transporte público 	Múltiple

Tipo de información	Formato de pregunta	Contenido	Formato de respuesta
		<ul style="list-style-type: none"> • 2 o más transportes públicos • Otro_____ 	
Lugar de accidentes grave que produjo incapacidad temporal	Multinomial	<ul style="list-style-type: none"> • Dentro de mi casa • En la calle, en mi día de descanso • En el trabajo o camino al trabajo • Ninguno, no sufrí accidentes graves 	Múltiple
Situación familiar grave en el último año	Multinomial	<ul style="list-style-type: none"> • Enfermedad • Deceso • Problemas económicos • Separaciones • Otros • Ninguna 	Múltiple
Actividades fuera del trabajo	Multinomial	<ul style="list-style-type: none"> • Deportes • Otro trabajo • Actividades religiosas • Descansar, ver TV • Tareas del hogar • Varias de las anteriores • Otras 	Múltiple
Incidente o accidente sin lesión grave	Ordinal	<ul style="list-style-type: none"> • Ninguno • Uno • Dos • Varios 	Única

Flujo de codificación

(instrucciones dadas al codificador)

1. **Leer manual de códigos.** Leer el documento donde se describen, definen y se muestran ejemplos específicos de los códigos. Es importante familiarizarse con estas descripciones.
2. **Escanear las respuestas.** En la base de datos, y empezando por el estresor 1 o motivador 1, leer todas las respuestas para familiarizarse con el contenido.
3. **Codificación.** Iniciar la codificación, asociando la respuesta al uno de los códigos existentes. Se hará una marca (X) en el casillero donde la respuesta parece estar asociada. En caso de dudas, se debe de regresar a leer el manual de códigos.
4. **Eficiencia de la codificación.** Asignar un 3, 2, o 1, en el casillero correspondiente a la respuesta, en la columna CALIFICACION, debajo de sobre el grado de eficiencia o rapidez para asignar el código. Estos números significan: *inmediata* (3; cuando no se percibió dudas sobre la clasificación y no requirió adicional examen del contenido), *moderadamente iterativa* (2; cuando se percibió moderada duda sobre la clasificación inicial), *intensamente iterativa* (1; cuando se percibió alta duda sobre la adecuada clasificación existente, y se requirió ser considerado como no clasificable en las categorías existentes
5. **Codificar en nuevas categorías.** Las respuestas que no fueron codificadas en las categorías iniciales, serán codificadas en nuevas categorías propuestas. Se tendrá que elegir alguna de estas nuevas categorías propuestas, asignando un "X" correspondiente con la respuesta analizada.

Códigos: descripción, etiqueta y definiciones

Demandas laborales

Definición. Es el conjunto de exigencias cognitivas y emocionales de la tarea, así como aquellas relacionadas al volumen de trabajo, velocidad, responsabilidad, peligros, del ambiente físico, horarios, jornadas y esfuerzos físicos en general que se imponen y se perciben por el trabajador en el contexto de la actividad laboral.

Código	Etiqueta	Definición (Juárez-García A. y Flores-Jiménez, 2020)	Descripción
1	Demandas por la carga y ritmo de trabajo	Exigencias percibidas respecto al monto de carga y ritmo del trabajo	Demasiada carga de trabajo, trabajar muy rápido, interrupciones que impiden completar las tareas a tiempo
2	Demandas por alta responsabilidad y peligrosidad	Exigencias percibidas respecto al grado de responsabilidad y peligrosidad implicadas en la realización del trabajo	Tomar acciones o decisiones de las que depende la vida o seguridad de otras personas, existencia de peligros para la integridad del trabajador o la de otros
3	Demandas por jornadas, turnos u horarios	Exigencias percibidas sobre la cantidad y cambios en las horas y horario de trabajo	Trabajar horas extras o en horarios prolongados, rotar o cambiar de turno, laborar jornadas en turno nocturno
4	Demandas cognitivas o atencionales	Exigencias percibidas sobre la intensidad de la atención y esfuerzo mental, y simultaneidad de tareas	Mantener la atención en las tareas por períodos prolongados, atender dos o más tareas al mismo tiempo, tareas con demasiado esfuerzo mental
5	Demandas emocionales (trato con personas)	Exigencias percibidas sobre las respuestas emocionales y el trato a personas (trabajadores y personas externas al trabajo)	Enfrentar emociones o estados de ánimo negativos de otras personas, trato con personas desgastante, mostrar emociones diferentes a las que siento
6	Demandas de esfuerzo físico	Exigencias percibidas sobre la cantidad, duración y posiciones implicados en el esfuerzo físico	Trabajo exigente de mucho esfuerzo físicamente, realizar trabajo en posiciones muy incómodas
7	Demandas por el ambiente físico	Características físicas laborales, que son deficientes o excesivas, y son adversas para realizar el trabajo	Trabajo en condiciones físico-ambientales deficientes o muy adversas (ya sea por excesivo ruido, polvos, temperatura, aire, espacio reducido, etc.)

ACOSO LABORAL

Definición. Es el maltrato verbal o modal que de manera frecuente y sistemática recibe un trabajador por parte de su jefe o compañeros

Código	Etiqueta	Definición (Juárez-García A. y Flores-Jiménez, 2020)	Descripción
8	Acoso psicológico por superiores	Estímulos o respuestas intencionadas de los superiores hostiles	Los superiores maltratan psicológicamente de forma intencional (ya sea que me humillan, bloquean, excluyen, amenazan, insultan, descalifican, se burlan, difaman o ignoran, etc.)
9	Acoso psicológico por compañeros o subordinados	Comportamiento intencionalmente hostil de los compañeros	Los compañeros o subordinados atacan o agreden psicológicamente de forma intencional (ya sea que me excluyen, amenazan, insultan, descalifican, se burlan, difaman o ignoran, etc.)

LIDERAZGO ESTRESANTE

Definición. Se refiere a un estilo de dirección de las personas caracteriza-do por una supervisión estricta, limitante de la autodeterminación y con una retroalimentación al desempeño ausente o negativa.

Código	Etiqueta	Definición (Juárez-García A. y Flores-Jiménez, 2020)	Descripción
10	Supervisión controlante	Monitoreo altamente controlante e inflexibilidad para cumplir las normas	Siento que en mi trabajo me quieren controlar demasiado, cumplimiento muy exagerado de reglas y normas que deben cumplir los empleados
11	Retroalimentación negativa o inadecuada del supervisor	Respuestas negativas sobre el desempeño laboral	Mis superiores informan y retroalimentan inadecuadamente sobre el desempeño en el trabajo

Recompensas y desarrollo laboral

Es el conjunto de beneficios y retribuciones que se reciben por motivo del trabajo tanto de carácter económico, como psicosocial y de desarrollo laboral.

Código	Etiqueta	Definición (Juárez-García A. y Flores-Jiménez, 2020)	Descripción
	Recompensas y desarrollo laboral	Es el conjunto de beneficios y retribuciones que se reciben por motivo del trabajo tanto de carácter económico, como psicosocial y de desarrollo laboral	Justicia organizacional, Salario motivante, Reconocimiento por el trabajo, Oportunidades para el desarrollo laboral y profesional, Seguridad en la conservación del empleo, tarea gratificante, tarea trascendente
	Control laboral	Es el grado de autonomía, claridad de funciones, libertad en el uso de habilidades y desarrollo de capacidades que permite un puesto de trabajo.	Autonomía, Utilización de habilidades, desarrollo de capacidades, Trabajo variado, Funciones y roles claros
	Recursos para realizar el trabajo	Se refiere a la disposición de materiales, equipos y herramientas físicas y de conocimiento que son necesarias para desempeñar de manera óptima el trabajo.	Recursos materiales, equipos y herramientas para el trabajo, Capacitación
	Clima y apoyo social en el trabajo	Se refiere a la percepción de integración social y de soporte psicológico de los compañeros y jefes en el trabajo.	Apoyo de compañeros, Apoyo de jefes, Clima de unión
	Congruencia de valores	Grado de ajuste y correspondencia entre los valores del trabajador y los de sus compañeros y de la empresa en general.	Congruencia de valores entre individuo-organización, y el individuo-compañeros-superiores

Estructura del formulario para la claridad y relevancia de los ítems

Formulario Para validez de contenido

Título de la investigación : Desarrollo y validación intercultural de un modelo de evaluación de riesgo psicosocial laboral
Institución : Universidad Autónoma del Estado de Morelos
Investigadores : César Merino Soto
Teléfonos : (52)777 4259409
Email : elsikander@yahoo.com.ar, camsresearch@yahoo.com

Presentación:

Estamos elaborando un cuestionario que evalúa **varios aspectos del trabajador y de su ambiente de trabajo**. Este cuestionario será aplicado a trabajadores hombres y mujeres, peruanos, de diferentes tipos de empleo y nivel de instrucción; el nivel de instrucción mínimo para comprender el contenido de este cuestionario es secundaria completa o incompleta.

Para asegurarnos que el cuestionario sea comprendido, le pedimos su colaboración para evaluar dos cosas. Primero, **la claridad de los ítems/afirmaciones**; esto se refiere al grado en que el ítem está redactado de una manera que es entendible por los participantes que responderán (trabajadores). Es importante que, cuando responda, únicamente piense en la claridad de la redacción del ítem, pero no en otros aspectos, como su interpretación, su relación con el desempleo, etc. Antes de calificar cada ítem, léalo completamente una o dos veces, y su impresión sobre la claridad del ítem lo registra marcando (con una "X") una de las opciones de respuesta, que va del **1 (Nada claro)** hasta **5 (Completamente claro)**; el valor intermedio es 4, que corresponde con una afirmación moderadamente comprensible (que no es tan claro, pero tampoco incomprensible):

Nada claro				Completamente claro
1	2	3	4	5

Segundo, le pediremos que, de acuerdo a su percepción, califique cada ítem/afirmación sobre el grado o **impacto de cada afirmación/ítem que produciría cuando es leído**; es decir, si al ser leído la afirmación por el encuestado, puede ocurrir en éste alguna reacción emocional (como enojo, ansiedad o depresión, preocupación, etc.) y/o una sensación de incomodidad general. Use una numeración del 1 (impacto trivial, realmente inofensivo) hasta 5 (impacto muy fuerte):

Ninguna Relevancia				Completa relevancia
1	2	3	4	5

Sección 1: DATOS GENERALES

Edad:	Género <input type="checkbox"/> Hombre <input type="checkbox"/> Mujer	¿Actualmente trabaja?
Últimos estudios (puede marcar varios)		Años de antigüedad en la profesión:
<input type="checkbox"/> Secundaria completa	<input type="checkbox"/> Completo <input type="checkbox"/> Incompleto	Años de antigüedad en el puesto de trabajo actual:
<input type="checkbox"/> Técnico	<input type="checkbox"/> Completo <input type="checkbox"/> Incompleto	
<input type="checkbox"/> Licenciatura	<input type="checkbox"/> Completo <input type="checkbox"/> Incompleto	Puesto de trabajo:
<input type="checkbox"/> Maestría	<input type="checkbox"/> Completo <input type="checkbox"/> Incompleto	
<input type="checkbox"/> Doctorado	<input type="checkbox"/> Completo <input type="checkbox"/> Incompleto	
<input type="checkbox"/> Segunda especialidad	<input type="checkbox"/> Completo <input type="checkbox"/> Incompleto	Carrera o profesión (si por ahora ninguna, deje en blanco):

Si desea información sobre los resultados de este estudio, puede comunicarse conmigo a la información de contacto en el encabezado de este formulario. Le agradezco su participación.

Sección 2: Preguntas previas:

Considerando el idioma castellano que tú manejas, ¿qué tan bueno eres..?				
... para entenderlo cuando te hablan?	Bajo	No tan bien	Bien	Muy bien
... para hablarlo (expresarte oralmente)?	Bajo	No tan bien	Bien	Muy bien
... para leerlo (comprender lo que lees)?	Bajo	No tan bien	Bien	Muy bien
... para escribirlo (expresarte por escrito)?	Bajo	No tan bien	Bien	Muy bien

¿Con qué frecuencia lees...?

Periódicos o revistas?	Muy poco	Algunas veces	Frecuente	A diario
Libros?	Muy poco	Algunas veces	Frecuente	A diario
Otros documentos (notas, reportes, etc.)?	Muy poco	Algunas veces	Frecuente	A diario

¿Con qué frecuencia necesita ayuda de otra persona cuando usted lee instrucciones, indicaciones, recetas u otro material escrito?	Nunca	Raramente	A veces	Frecuente	Siempre
¿En qué grado puede llenar bien formularios, cuestionarios, encuestas, etc.?	Nada bien	Un poco bien	Más o menos bien	Bien	Muy bien
¿Con qué frecuencia tiene problemas para comprender información o instrucciones escritas?	Siempre	Frecuente	A veces	Raramente	Nunca

Sección 3: Claridad y Relevancia:

Contenido del ítem/afirmación	Relevancia					Claridad				
	Nada relevante		Completa-mente relevante			Nada claro		Completa-mente claro		
1. Tengo demasiada carga de trabajo	1	2	3	4	5	1	2	3	4	5
2. Tengo que trabajar muy rápido	1	2	3	4	5	1	2	3	4	5
3. En mi trabajo tengo interrupciones que me impiden completar mis tareas a tiempo	1	2	3	4	5	1	2	3	4	5
4. Mi trabajo exige tomar acciones o decisiones de las que depende la vida o seguridad de otras personas	1	2	3	4	5	1	2	3	4	5
5. En mi trabajo existen peligros para mi integridad o la de otros	1	2	3	4	5	1	2	3	4	5
6. Debo trabajar horas extras o en horarios que se prolongan mucho	1	2	3	4	5	1	2	3	4	5
7. Debo rotar o cambiar de turno	1	2	3	4	5	1	2	3	4	5
8. Debo laborar jornadas en turno nocturno	1	2	3	4	5	1	2	3	4	5
9. En mi trabajo tengo que mantener la atención en mis tareas por períodos prolongados	1	2	3	4	5	1	2	3	4	5
10. En mi trabajo tengo que atender dos o más tareas al mismo tiempo.	1	2	3	4	5	1	2	3	4	5
11. Mi trabajo requiere demasiado esfuerzo mental	1	2	3	4	5	1	2	3	4	5
12. En mi trabajo debo enfrentarme con emociones o estados de ánimo negativos de otras personas	1	2	3	4	5	1	2	3	4	5
13. El trato con personas en mi trabajo es desgastante	1	2	3	4	5	1	2	3	4	5
14. Mi trabajo me exige mostrar emociones diferentes a las que siento	1	2	3	4	5	1	2	3	4	5
15. En mi trabajo tengo que esforzarme mucho físicamente	1	2	3	4	5	1	2	3	4	5
16. Tengo que realizar mi trabajo en posiciones muy incómodas	1	2	3	4	5	1	2	3	4	5
17. En mi trabajo existen condiciones físico- ambientales deficientes o muy adversas (ya sea por excesivo ruido, polvos, temperatura, aire, espacio reducido, etc.)	1	2	3	4	5	1	2	3	4	5
18. En mi trabajo <u>mis superiores</u> me maltratan psicológicamente de forma intencional (ya sea que me humillan, bloquean, excluyen, amenazan, insultan, descalifican, se burlan, difaman o ignoran, etc.)	1	2	3	4	5	1	2	3	4	5
19. En mi trabajo <u>mis compañeros o subordinados</u> me maltratan psicológicamente de forma intencional (ya sea que me humillan, bloquean, excluyen, amenazan, insultan, descalifican, se burlan, difaman o ignoran, etc.)	1	2	3	4	5	1	2	3	4	5
20. Siento que en mi trabajo me quieren controlar demasiado	1	2	3	4	5	1	2	3	4	5
21. Creo que las reglas y normas que deben cumplir los empleados en esta organización son muy exageradas	1	2	3	4	5	1	2	3	4	5
22. Mis superiores me informan y retroalimentan adecuadamente sobre mi desempeño en el trabajo	1	2	3	4	5	1	2	3	4	5
23. En mi trabajo las cosas se manejan de manera justa y equitativa.	1	2	3	4	5	1	2	3	4	5
24. El salario que recibo me mantiene motivado	1	2	3	4	5	1	2	3	4	5
25. Mi trabajo es bien apreciado y reconocido por muchas personas	1	2	3	4	5	1	2	3	4	5
26. Dudo que existan oportunidades de crecer laboral o profesionalmente en mi trabajo	1	2	3	4	5	1	2	3	4	5
27. Pienso que puedo perder este empleo próximamente	1	2	3	4	5	1	2	3	4	5
28. Mi trabajo es muy agradable y gratificante	1	2	3	4	5	1	2	3	4	5
29. Las tareas que realizo en mi trabajo benefician de forma importante a otras personas y a la sociedad	1	2	3	4	5	1	2	3	4	5
30. Soy libre en decidir cómo hacer mi trabajo	1	2	3	4	5	1	2	3	4	5
31. En mi trabajo utilizo las habilidades en las que soy bueno(a)	1	2	3	4	5	1	2	3	4	5

Contenido del ítem/afirmación	Relevancia					Claridad				
	Nada relevante		Completamente relevante			Nada claro		Completamente claro		
32. Mi trabajo me permite desarrollar mis capacidades	1	2	3	4	5	1	2	3	4	5
33. Las actividades que realizo en mi trabajo son muy variadas	1	2	3	4	5	1	2	3	4	5
34. Son perfectamente claras las funciones y tareas que me corresponden en mi trabajo	1	2	3	4	5	1	2	3	4	5
35. Cuento con los materiales, equipos y herramientas necesarias y adecuadas para hacer mi trabajo	1	2	3	4	5	1	2	3	4	5
36. Recibo las capacitaciones necesarias para realizar el trabajo de manera efectiva	1	2	3	4	5	1	2	3	4	5
37. Mis compañeros de trabajo me apoyan en situaciones difíciles	1	2	3	4	5	1	2	3	4	5
38. Mi jefe o mis superiores me apoyan en situaciones difíciles	1	2	3	4	5	1	2	3	4	5
39. Existe un clima de unión y colaboración en mi trabajo	1	2	3	4	5	1	2	3	4	5
40. Encuentro grandes coincidencias entre mis valores y los de la organización o lugar donde trabajo	1	2	3	4	5	1	2	3	4	5
41. Los valores de mis compañeros de trabajo y mis superiores coinciden totalmente con los míos	1	2	3	4	5	1	2	3	4	5

Instrumento: PROPSIT (Procesos Psicosociales en el Trabajo)

**ENCUESTA DE TAMIZAJE GENERAL
PROPSIT (PROCESOS PSICOSOCIALES EN EL TRABAJO)**

Para responder a este cuestionario, considere que no hay respuestas buenas ni malas. Es importante que conteste a **todas** las frases de una manera **sincera**; por favor, no deje alguna sin contestar.

Use la siguiente numeración para responder:

0	1	2	3	4	5	6
Nunca	Rara vez (muy pocas veces al año o menos)	De vez en cuando (Una vez al mes o menos)	Regularmente (Algunas veces al mes)	Frecuentemente (Una vez por semana)	Muy Frecuentemente (Varias veces por semana)	Siempre: (Todos los días)

Marque el número que corresponda para indicar con qué frecuencia ha experimentado estos sentimientos o situaciones. Si nunca lo ha experimentado, marque 0. Si en efecto ha experimentado este pensamiento o sentimiento, marque la respuesta que mejor lo describa.

Parte 1

	Nunca	Rara vez	De vez en cuando	Regularmente	Frecuentemente	Muy frecuentemente	Siempre
	0	1	2	3	4	5	6
1. Tengo demasiada carga de trabajo	0	1	2	3	4	5	6
2. Tengo que trabajar muy rápido	0	1	2	3	4	5	6
3. En mi trabajo tengo interrupciones que me impiden completar mis tareas a tiempo	0	1	2	3	4	5	6
4. Mi trabajo exige tomar acciones o decisiones de las que depende la vida o seguridad de otras personas	0	1	2	3	4	5	6
5. En mi trabajo existen peligros para mi integridad o la de otros	0	1	2	3	4	5	6
6. Debo trabajar horas extras o en horarios que se prolongan mucho	0	1	2	3	4	5	6
7. Debo rotar o cambiar de turno	0	1	2	3	4	5	6
8. Debo laborar jornadas en turno nocturno	0	1	2	3	4	5	6
9. En mi trabajo tengo que mantener la atención en mis tareas por períodos prolongados	0	1	2	3	4	5	6
10. En mi trabajo tengo que atender dos o más tareas al mismo tiempo.	0	1	2	3	4	5	6
11. Mi trabajo requiere demasiado esfuerzo mental	0	1	2	3	4	5	6
12. En mi trabajo debo enfrentarme con emociones o estados de ánimo negativos de otras personas	0	1	2	3	4	5	6
13. El trato con personas en mi trabajo es desgastante	0	1	2	3	4	5	6
14. Mi trabajo me exige mostrar emociones diferentes a las que siento	0	1	2	3	4	5	6
15. En mi trabajo tengo que esforzarme mucho físicamente	0	1	2	3	4	5	6
16. Tengo que realizar mi trabajo en posiciones muy incómodas	0	1	2	3	4	5	6
17. En mi trabajo existen condiciones físico- ambientales deficientes o muy adversas (ya sea por excesivo ruido, polvos, temperatura, aire, espacio reducido, etc.)	0	1	2	3	4	5	6
18. En mi trabajo <u>mis superiores</u> me maltratan psicológicamente de forma intencional (ya sea que me humillan, bloquean, excluyen, amenazan, insultan, descalifican, se burlan, difaman o ignoran, etc.)	0	1	2	3	4	5	6
19. En mi trabajo <u>mis compañeros o subordinados</u> me maltratan psicológicamente de forma intencional (ya sea que me humillan, bloquean, excluyen, amenazan, insultan, descalifican, se burlan, difaman o ignoran, etc.)	0	1	2	3	4	5	6

							Siempre
						Muy frecuentemente	
						Frecuentemente	
					Regularmente		
		De vez en cuando					
		Rara vez					
	Nunca						

20. Siento que en mi trabajo me quieren controlar demasiado	0	1	2	3	4	5	6
21. Creo que las reglas y normas que deben cumplir los empleados en esta organización son muy exageradas	0	1	2	3	4	5	6
22. Mis superiores me informan y retroalimentan adecuadamente sobre mi desempeño en el trabajo	0	1	2	3	4	5	6

23. En mi trabajo las cosas se manejan de manera justa y equitativa.	0	1	2	3	4	5	6
24. El salario que recibo me mantiene motivado	0	1	2	3	4	5	6
25. Mi trabajo es bien apreciado y reconocido por muchas personas	0	1	2	3	4	5	6
26. Dudo que existan oportunidades de crecer laboral o profesionalmente en mi trabajo	0	1	2	3	4	5	6
27. Pienso que puedo perder este empleo próximamente	0	1	2	3	4	5	6
28. Mi trabajo es muy agradable y gratificante	0	1	2	3	4	5	6
29. Las tareas que realizo en mi trabajo benefician de forma importante a otras personas y a la sociedad	0	1	2	3	4	5	6

30. Soy libre en decidir cómo hacer mi trabajo	0	1	2	3	4	5	6
31. En mi trabajo utilizo las habilidades en las que soy bueno(a)	0	1	2	3	4	5	6
32. Mi trabajo me permite desarrollar mis capacidades	0	1	2	3	4	5	6
33. Las actividades que realizo en mi trabajo son muy variadas	0	1	2	3	4	5	6
34. Son perfectamente claras las funciones y tareas que me corresponden en mi trabajo	0	1	2	3	4	5	6

35. Cuento los materiales, equipos y herramientas necesarias y adecuadas para hacer mi trabajo	0	1	2	3	4	5	6
36. Recibo las capacitaciones necesarias para realizar el trabajo de manera efectiva	0	1	2	3	4	5	6

37. Mis compañeros de trabajo me apoyan en situaciones difíciles	0	1	2	3	4	5	6
38. Mi jefe o mis superiores me apoyan en situaciones difíciles	0	1	2	3	4	5	6
39. Existe un clima de unión y colaboración en mi trabajo	0	1	2	3	4	5	6

40. Encuentro grandes coincidencias entre mis valores y los de la organización o lugar donde trabajo	0	1	2	3	4	5	6
41. Los valores de mis compañeros de trabajo y mis superiores coinciden totalmente con los míos	0	1	2	3	4	5	6

Parte 2: Por favor, responda a estas preguntas

¿Con qué frecuencia hace ejercicio?	0	1	2	3	4	5	6
¿Con qué frecuencia fuma?	0	1	2	3	4	5	6
¿Con qué frecuencia toma bebidas alcohólicas?	0	1	2	3	4	5	6
¿Con qué frecuencia toma alguna droga o estupefaciente?	0	1	2	3	4	5	6
¿Con qué frecuencia toma alguna bebida energizante?	0	1	2	3	4	5	6
¿Con qué frecuencia consume alimentos que Ud sabe que dañan su salud?	0	1	2	3	4	5	6

Instrumentos adicionales aplicados

UWES-3 (Utrecht Work Engagement Scale – 3 items)

A continuación, hay algunas afirmaciones respecto a los sentimientos que le provocan su trabajo

	Siempre						
	Muy frecuentemente						
	Frecuentemente						
	Regularmente						
	De vez en cuando						
	Rara vez						
	Nunca						
	0	1	2	3	4	5	6
1. En mi trabajo me siento lleno de energía	0	1	2	3	4	5	6
2. Estoy entusiasmado sobre mi trabajo	0	1	2	3	4	5	6
3. Estoy inmerso en mi trabajo	0	1	2	3	4	5	6

f. Medida de estrés

SOS – S

Instrucciones

En las siguientes páginas, hallará afirmaciones sobre cómo se sentía usted durante la **semana pasada**. Por favor, responda a cada afirmación, aun cuando parezcan similares. Cada afirmación nombra determinados sentimientos que lo experimentan en su vida diaria. Marque su respuesta a cada afirmación en uno de los siguientes niveles:

Nada					Mucho
1	2	3	4	5	

Por ejemplo, usted podría marcar la última marca (5) si hubiera sentido una fuerte intensidad; o podría hacer una X en la marca si hubiera sentido solo un poco. Sea tan honesto como sea posible. Sus respuestas no ayudarán a comprender las emociones en la vida diaria. Sus respuestas serán confidenciales. No necesita escribir su nombre o algo que lo identifique en este cuestionario. Por favor, recuerde considerar únicamente sus sentimientos en la **semana pasada**.

En la semana pasada, usted sintió que...						
1 que las cosas no estaban como esperaba	Nada				Mucho
		1	2	3	4	5
2	... que estoy abrumado por mis responsabilidades	Nada				Mucho
		1	2	3	4	5
3	... que las oportunidades estaban en mi contra	Nada				Mucho
		1	2	3	4	5
4 que no había suficiente tiempo para hacer las cosas	Nada				Mucho
		1	2	3	4	5
5	... como si nada estuviera bien	Nada				Mucho
		1	2	3	4	5
6	... como si estuviera apurado	Nada				Mucho
		1	2	3	4	5
7 como si no hubiera salida o escape	Nada				Mucho
		1	2	3	4	5
8	... como si las cosas se estuvieran amontonando	Nada				Mucho
		1	2	3	4	5
9	... como si quisiera darse por vencido	Nada				Mucho
		1	2	3	4	5
10	... como si estuviera llevando una carga pesada	Nada				Mucho
		1	2	3	4	5
11	... incapaz de enfrentar las cosas	Nada				Mucho
		1	2	3	4	5

g. Medidas de ítems únicos

Estrés

Estrés significa una situación en que una persona se siente tensa, inquieta, nerviosa o ansiosa, que tiene problemas para dormir, debido a su mente está preocupada todo el tiempo. ¿Usted se siente así en estos días?

- (1) Casi nada
- (2)
- (3)
- (4)
- (5) Mucho

Cuernavaca, Mor; a 21 de **Octubre** de 2022

DR. EDUARDO HERNÁNDEZ PADILLA
COORDINADOR ACADÉMICO DEL DOCTORADO EN PSICOLOGÍA
P R E S E N T E

Por este medio, me permito informar a usted el dictamen de los **votos aprobatorios** de la revisión de la tesis titulada: **Validación de un modelo de evaluación de riesgo psicosocial laboral, en trabajadores peruanos y mexicanos**, trabajo que presenta el **Mtro. César Merino Soto**, quien cursó el **Doctorado En Psicología** en el Centro De Investigación Transdisciplinar En Psicología (CITPsi) de la UAEM.

Sirva lo anterior para que dicho dictamen permita realizar los trámites administrativos correspondientes para la presentación de su examen de grado.

A T E N T A M E N T E

VOTOS APROBATORIOS			
COMISIÓN REVISORA	APROBADO	CONDICIONADA A QUE SE MODIFIQUEN ALGUNOS ASPECTOS*	SE RECHAZA*
Dr. Luis Fernando Arias Galicia	Aprobado		
Dr. Sergio Alexis Domínguez Lara	Aprobado		
Dr. José Manuel Fernández Arata	Aprobado		
Dr. Pedro Rafael Gil Monte	Aprobado		
Dr. Eduardo Hernández Padilla	Aprobado		
Dr. Arturo Juárez García	Aprobado		
Dr. Luis Pérez Álvarez	Aprobado		

*En estos casos deberá notificar al/la estudiante el plazo dentro del cual deberá presentar las modificaciones o la nueva investigación (no mayor a 30 días).



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS

Se expide el presente documento firmado electrónicamente de conformidad con el ACUERDO GENERAL PARA LA CONTINUIDAD DEL FUNCIONAMIENTO DE LA UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MORELOS DURANTE LA EMERGENCIA SANITARIA PROVOCADA POR EL VIRUS SARS-COV2 (COVID-19) emitido el 27 de abril del 2020.

El presente documento cuenta con la firma electrónica UAEM del funcionario universitario competente, amparada por un certificado vigente a la fecha de su elaboración y es válido de conformidad con los LINEAMIENTOS EN MATERIA DE FIRMA ELECTRÓNICA PARA LA UNIVERSIDAD AUTÓNOMA DE ESTADO DE MORELOS emitidos el 13 de noviembre del 2019 mediante circular No. 32.

Sello electrónico

SERGIO ALEXIS DOMINGUEZ LARA | Fecha:2022-10-21 10:25:54 | Firmante

XNQixL4LzCxp4mxOXWE+dolrFCE1dRrMaKr7MHu0FCGZgk02D0eo+LC7aPSzUogc6b9PNTbYcJrN99b9/IKHpywc7JIOcmz+t871OfWC4HJ/VJsbjMC7jZD1k93GlxHwKDemSg+WLLRMVYdMMa5Fi+cp30J2/bWfW8rkAK2ee+iy0CUQTf3i91YgUFBpCsgGureVwodNU7u7RcKpbZiq3aGhBr1cHsCfsRt4aSG6VRKZGeTC1CRDQ6fdLit29mtvjz2hQYqK65mD6AuNthDy/QgZ19UMTrRgHPqikK+7UJIBQQSsx7jX2ilVWtpxkeAkWlvzITqWlmqS5zhZXEg==

EDUARDO HERNANDEZ PADILLA | Fecha:2022-10-21 10:40:34 | Firmante

Ds7JBYPgZYoscG54i0cXumtFRH3ZmLhxBtFuiC0sb4LL0879PWhuiCARsVbu89r/Xb/FoGmQcdqWlotQ04EP4AW9/JQNNmXvWldtF7AKTJgiHdtFyazolQTnn5HT/ZFR4dN61f09XPNoHg02RXDqjijKr3I56nmQDeZiR0Cr92ktOHpWjjZx9iF0bcQyO+wEQNh7f8qzUKVE+eKOK6/cN2yCz94eS7+dt5cpCRnQ5zyuJRa7FhpqQlqysQ03METc0EYPDzQZ11A57hVXK8hyBhikOyRnCVlbu0WeUFuQo1PAhyWmHYT/Ugsd5ad2wiAAwWMzEXxaoJhghsQ==

ARTURO JUAREZ GARCIA | Fecha:2022-10-21 10:46:58 | Firmante

bDEYEIOsu+o8I3V/ReUep+HmoQ/y5mUCTr9W7RQJwLoPSxammsijvlyNwL+/j7U2uOsCYKIJY8+OGzTcR2zKOhza3cG1lwdh9+Scv72ObjeFalSEfjwManzPaK6PcyGVLFWaoP/bYbaq2pmNW4LDRykqG6ySJS5gaiq2z1UZmWUNgxVcbOh13LgJVvwoRGY56DrQ/gerVo1xHD9gp47bbakLQLFvz2NP4VTCUXpxCq+Qc6Ns8atPOXiJdug+DZDMLhU6TLV OENCZ5cvv/o8Dzdzsf1MOE9hQeXdvZ+zBLsr+ZlcYX70RjMLTfMRRejeMRQmFELBvmpRjiAR0Gnw==

JOSE MANUEL FERNANDEZ ARATA | Fecha:2022-10-21 11:07:26 | Firmante

QU4L3o6NxfU8zKyhXij8JqZWs/Ks5KGyZ1wYFBAoA5aiYaaMLmWCZn2BpbZxBRFYNOplTS9hasbi1fZAZkf5XCFMO8bMWqzdSH1P5IDxvzuVA1VPm243GEIwhpP4EZb0otEYgPUWKf+PuHMKrTbGR2MPJToZfsiULmS9RatIYi38kFqaQv7c6GKhPZxjUtAXLIGmLTSLOycqlosYwStmFoPIVdNdRj97b4un8wC+Lq745pynOAOXvwnzP0IY7doUgEqvF+Vv0B6R86wGBhp/qO9dlullfptaoZ2WHez+Gu618QE3T9M2feyMTDzNvGJZNy9WdPE1zGia+Edl8A==

PEDRO RAFAEL GIL MONTE | Fecha:2022-10-21 13:14:35 | Firmante

I7ECWgnyVklgUzNS5ufLI0EHETnH9QMNFZokfyjhRjKxvFBle2VehKWE64sMT04EHDV15i2phkKZEEdYBn5Vb2M7mKXN2U6z8sdgksdiO9dvxcviEQK89Z2Rjvt79nzvCBFRJ8DPrV8JeNn5FWLTVbdGcGIRmjSSiTDzdlGgcDfOMw1buZm+z4a4K2he84XM234JO8zX/qx2ThU17iO5cDA3114varOZlbdjNk2i7+juBwb8nVYT3yQnzi404cWN8dlcr8fvlpZdq2Ya gEowjhlKdJUMD/yeUo7yFPgp1ePxpLI+SOMfeQSDg2Y9SS7SzSy8sEtG9e46sl1wCg==

LUIS PEREZ ALVAREZ | Fecha:2022-10-21 20:30:33 | Firmante

ndKLwKUDuU1/fu4q9jimVcRvo0T2HTsNRT3rdwqODOHGiNj3GbzGwFnQyJ7qnmztOfzHKf6xw05RooodpCB6LbgRj3Bv8q1JZXDA8VNJdw5oYzb7lep6eF0NgHjbc3WZgdnZ36vdwGyNtoFGNSRt92wnhnGjWaO/RcZXRrhqkl9PEVwYwaeG++jbA3z0XbKAolefi8vWhrz/fPAaoX4vGZHyLcQeDS19Aj842w5hg8dBy95/M6aoL6O9xockPyWjriMgD9Dbhak3WWCvNyXbRbS9xR1JTI4O53p/urgWlW9HSxHN8wC11aicS5FQ26iGTFDhTcP/Sx34VphlZg==

LUIS FERNANDO ARIAS GALICIA | Fecha:2022-11-09 13:21:27 | Firmante

D81+viXTPEzHDUiriGcmHJxTmR1mH7c8UNp9lWzgwWDPuqB9qei2MZ+b45HJg19G9OfarUuigftxsrKZTBcroAaJstjFy+J96tQC9WjTEO3mypi3HvSFTTYypjf1a6SSHsqbP4WXizPw+4SwDjfhVwri1SQS969S+cXcZXCBeMXIG2PU8xULLuO/gV/VJgK0kix1g8waxbM6iyJSGS+i4kfrvm+PszFlyMXGx703FFSHXn3leKAm8Ohr+PpigbTOGy4xnUbCtk2Q0thNC/8onFsljyAsj9QB1tAvY3aCFU+ZnmDba+kMJ+4lnqP8Nie0Fw6kMHYB9yr5md6DDP76Q==

Puede verificar la autenticidad del documento en la siguiente dirección electrónica o escaneando el código QR ingresando la siguiente clave:



EMlbHAVcC

<https://efirma.uaem.mx/noRepudio/PMWULUQj6U5O6U0mMUkepcCqz6D73j4>

