



BRECHAS DE PREPARACIÓN DIGITAL E INFORMACIONAL EN LA EMPRESA LATINOAMERICANA: UNA EVALUACIÓN MULTIPAÍS

DIGITAL AND INFORMATION READINESS GAPS IN LATIN AMERICAN COMPANIES: A MULTI-COUNTRY ASSESSMENT

José Isaac Zablah

Universidad Nacional Autónoma de Honduras, Honduras

mrzablah@yahoo.com

<https://orcid.org/0000-0002-1410-7189>

Lupe Carolina Espinoza

Universidad Técnica Particular de Loja, Ecuador

icespinoza@utpl.edu.ec

<https://orcid.org/0000-0003-1486-5622>

Lilian Elisa Sosa Díaz

Universidad Nacional Autónoma de Honduras, Honduras

liliansosa2012@gmail.com

<https://orcid.org/0000-0001-8148-9482>

Recibido: 27 enero del 2026

Revisado: 19 marzo de 2026

Aprobado: 2 de junio del 2026

Cómo citar: Isaac Zablah, J.; Carolina Espinoza, L. y Sosa, L. (2026). Brechas de preparación digital e informacional en la empresa Latinoamericana: una evaluación multipaís. *Bibliotecas, Anales de Investigación*, 22, Especial, 1-23

RESUMEN

Objetivo. Examinar si la preparación digital y la capacidad operativa de gestión de la información se desarrollan de forma paralela en las empresas latinoamericanas de países determinados, e identificar las características organizativas asociadas a configuraciones equilibradas y asimétricas de preparación. **Diseño/Metodología/Enfoque.** Se armonizaron microdatos de 2143 establecimientos con veinte trabajadores o más de Colombia, Costa Rica, El Salvador, México, Paraguay y Perú, procedentes de las Encuestas de Empresas del Banco Mundial de 2023. Se aplicó un análisis de conglomerados ponderado y se interpretaron los modelos de clasificación mediante valores de Shapley. La estructura dimensional de ambas medidas se contrastó mediante análisis factorial confirmatorio sobre matrices policóricas. **Resultados/Discusión.** Ambas dimensiones se asociaron de manera positiva, con análisis correlacionales, El perfil digitalmente avanzado

persistió sin los ceros estructurales, si bien con un centroide informacional muy superior (38,67 frente a 10,09). El contexto nacional, el tamaño de la empresa, la capacitación formal y la certificación de calidad fueron los factores explicativos de mayor peso, si bien la discriminación predictiva resultó moderada (exactitud balanceada de 0,49 a 0,58). **Conclusiones.** La adopción digital no supone necesariamente un uso maduro de la información en la operación, y las configuraciones de preparación conservan un carácter marcadamente nacional. La brecha propuesta es un contraste descriptivo entre dominios y no debe emplearse como índice de ordenamiento entre países. **Originalidad/Valor.** El estudio propone un marco conjunto y reproducible para detectar el desajuste entre la capacidad digital y la gestión basada en información, a partir de microdatos comparables a escala de empresa y aprendizaje automático explicable.

PALABRAS CLAVE: preparación digital; gestión de la información; preparación organizativa; América Latina; aprendizaje automático explicable; desacoplamiento medios-fines; encuestas de empresas

ABSTRACT

Objective. To examine whether digital readiness and operational information management capacity develop in parallel within Latin American companies across specific countries, and to identify the organizational characteristics associated with balanced and asymmetric readiness configurations.

Design/Methodology/Approach. Microdata from 2,143 establishments with twenty or more employees—sourced from the 2023 World Bank Enterprise Surveys across Colombia, Costa Rica, El Salvador, Mexico, Paraguay, and Peru—were harmonized. A weighted cluster analysis was applied, and classification models were interpreted using Shapley values. The dimensional structure of both measures was validated via confirmatory factor analysis based on polychoric matrices. **Results/Discussion.** Both dimensions were positively associated according to correlational analyses. The "digitally advanced" profile persisted even when excluding structural zeros, albeit with a significantly higher informational centroid (38.67 vs. 10.09). National context, firm size, formal training, and quality certification were the most significant explanatory factors, although predictive discrimination was moderate (balanced accuracy ranging from 0.49 to 0.58). **Conclusions.** Digital adoption does not necessarily imply mature operational use of information, and readiness configurations remain strongly influenced by national context. The proposed gap represents a descriptive contrast between domains and should not be used as a ranking index for comparing countries. **Originality/Value.** This study proposes a joint, reproducible framework to detect the misalignment between digital capacity and information-based management, utilizing comparable firm-level microdata and explainable machine learning.

KEYWORDS: digital readiness; information management; organizational readiness; Latin America; explainable machine learning; means-ends decoupling; enterprise surveys

INTRODUCCIÓN

La transformación digital ocupa hoy un lugar central en la competitividad empresarial, pero el valor organizativo de las tecnologías digitales no se agota en su adopción. La investigación reciente entiende la transformación digital como un proceso que articula activos tecnológicos con respuestas estratégicas, estructuras organizativas, capacidades y cambios en la creación de valor, y no como una simple acumulación de tecnologías de la información (Vial, 2019; Verhoef et al., 2021). La distinción resulta especialmente pertinente en América Latina y el Caribe, donde la digitalización se incorpora cada vez con mayor fuerza a las políticas dirigidas a la pequeña y mediana empresa, mientras las diferencias persistentes en capacidad productiva, competencias, financiamiento y condiciones institucionales siguen condicionando la posibilidad de explotar los recursos digitales (OCDE/CAF/SELA, 2024; Comisión Económica para América Latina y el Caribe [CEPAL], 2025).

La evidencia disponible sobre empresas latinoamericanas ya indica que la madurez digital es heterogénea y que tanto la capacitación como la experiencia digital acumulada pueden condicionar el desarrollo empresarial. González-Tamayo et al. (2023), con datos de pequeñas y medianas empresas de cinco países de la región, documentaron que las competencias y la experiencia digitales constituyen componentes relevantes de la relación entre madurez digital y desarrollo de la firma. El diagnóstico regional sobre transformación digital de las micro, pequeñas y medianas empresas añade una precisión que interesa directamente a este trabajo: la penetración de las tecnologías digitales se estratifica según el tamaño de la firma y el grado de madurez de la

tecnología considerada, y la brecha se ensancha justamente al pasar de la conectividad básica hacia los usos que exigen reorganización interna (Dini et al., 2021). Tales hallazgos concuerdan con una literatura más amplia según la cual la estrategia digital rinde por debajo de lo esperado cuando las capacidades de implementación, las rutinas organizativas y los arreglos directivos no evolucionan al ritmo de la adopción tecnológica (Correani et al., 2020). No conviene, por tanto, suponer que la existencia de sitios web, conectividad o transacciones electrónicas conlleva una capacidad equivalente de convertir la información en acción directiva estructurada.

La capacidad de gestión de la información aporta la perspectiva organizativa complementaria. Mithas et al. (2011) mostraron que esta capacidad sostiene las capacidades de gestión de clientes, de procesos y del desempeño; trabajos posteriores subrayaron que el éxito en la gestión de la información depende del compromiso directivo sostenido y de las rutinas organizativas, más que de la tecnología por sí misma (Kettinger et al., 2021). La preparación empresarial en gestión de la información se ha caracterizado asimismo como la aptitud para afrontar los desafíos de calidad, gobernanza, integración, ciclo de vida y uso estratégico de la información (Hausmann et al., 2014). En el plano operativo, prácticas como el seguimiento de indicadores de desempeño, la definición de metas explícitas, el empleo de horizontes temporales apropiados y la difusión interna de esas metas configuran una arquitectura práctica de información para el control y el aprendizaje directivos.

La distinción entre capacidad tecnológica e informacional también se aprecia en la literatura sobre prácticas de gestión. Las prácticas estructuradas de gestión difieren de manera sustancial entre empresas y se asocian con el desempeño organizativo, incluso dentro de un mismo entorno institucional (Bloom y Van Reenen, 2007; Bloom et al., 2019). Las Encuestas de Empresas del Banco Mundial adaptan esa tradición al medir el seguimiento, la fijación de metas y las prácticas de personal mediante un instrumento internacional armonizado. Los análisis más recientes basados en esos datos siguen revelando una dispersión considerable de las prácticas de gestión entre países y entre empresas (Viganola y Diallo, 2025). Para las empresas latinoamericanas, ello abre la posibilidad de examinar la capacidad operativa digital y la capacidad directiva basada en información de forma conjunta, en lugar de tratarlas como manifestaciones intercambiables de la modernización organizativa.

La investigación sobre preparación organizativa se ha expandido con rapidez al compás de la difusión de la inteligencia artificial. Marcos recientes evalúan el grado de alistamiento de las organizaciones para adoptar inteligencia artificial en sus dimensiones tecnológica, organizativa, ambiental y humana (Kalleparambil y Akoum, 2025; Naheed et al., 2025). Los métodos de aprendizaje automático explicable se han empleado además para perfilar la preparación organizativa y exponer las variables asociadas a los estados organizativos predichos (Shafiabady et al., 2024; Sriwichien et al., 2026). En el plano regional, las evaluaciones de preparación digital construidas por agregación de índices internacionales ordenan a los países latinoamericanos en estratos de desarrollo y sitúan a las seis economías aquí analizadas en posiciones contrastantes, pero operan sobre agregados nacionales y no permiten observar si la capacidad digital y la gestión informacional coexisten dentro de una misma firma (Chavarry Galvez y Revinova, 2025). Con todo, la revisión no sistemática de la literatura realizada para este estudio no localizó un análisis multipaís a escala de empresa en América Latina que modele explícitamente la alineación entre la capacidad operativa digital observable y la preparación operativa en gestión de la información, y que emplee después el aprendizaje automático explicable como capa interpretativa y no como el propio constructo de preparación. Esta afirmación sobre el vacío existente descansa en una búsqueda dirigida más que exhaustiva, y se ofrece como motivación del diseño propuesto antes que como reivindicación de novedad absoluta.

El presente estudio evalúa, en consecuencia, la configuración conjunta de la preparación digital y de la preparación operativa en gestión de la información entre las empresas medianas y grandes de seis países latinoamericanos. El objetivo consiste en determinar el grado de alineación entre ambas dimensiones, identificar perfiles de preparación con respaldo empírico, cuantificar las diferencias entre países y examinar qué características externas de la firma distinguen los perfiles resultantes. El análisis introduce una Brecha estandarizada de Preparación Digital-Informacional como contraste descriptivo entre dominios y combina estimación ponderada por diseño muestral, validación de conglomerados, clasificación supervisada e interpretación de modelos mediante valores de Shapley.

Cuatro preguntas de investigación ordenan el trabajo. Primera: ¿en qué grado se alinean, a escala de establecimiento, la capacidad operativa digital y la preparación operativa en gestión de la información? Segunda: ¿qué configuraciones conjuntas admiten respaldo empírico reproducible y, en particular, emerge un perfil que combine capacidad informacional con restricción digital? Tercera: ¿qué magnitud alcanzan las diferencias entre países una vez incorporada la estructura de ponderación del diseño muestral? Cuarta: ¿qué características organizativas ajenas a la construcción de ambos índices distinguen los perfiles resultantes, y supera ese poder discriminante el que cabría esperar por azar?

METODOLOGÍA

Fuente de datos y muestra analítica

Se utilizaron microdatos a nivel de establecimiento de las Encuestas de Empresas del Banco Mundial de 2023 correspondientes a Colombia, Costa Rica, El Salvador, México, Paraguay y Perú (Grupo Banco Mundial, 2025). Estas encuestas aplican un cuestionario estandarizado y un muestreo aleatorio estratificado, con el sector de actividad, el tamaño de la empresa y la localización geográfica como estratos principales, lo que permite que las estimaciones nacionales reflejen la población de establecimientos elegibles del sector privado formal cuando se aplican los ponderadores muestrales (Banco Mundial, 2026). Los seis archivos de origen contenían 4692 establecimientos antes de las restricciones analíticas.

El Banco Mundial define los indicadores de prácticas de gestión para establecimientos con veinte empleados o más. Por ello se recurrió a la variable de tamizaje de empleo para restringir la muestra a esa población objetivo y se exigió información completa en los cuatro componentes del Índice de Preparación Digital y en los cuatro componentes de Preparación Operativa en Gestión de la Información. La muestra primaria final quedó conformada por 2143 establecimientos: México ($n = 682$), Colombia ($n = 443$), Perú ($n = 364$), El Salvador ($n = 285$), Paraguay ($n = 192$) y Costa Rica ($n = 177$). Las estimaciones nacionales emplearon la variable de la encuesta denominada «ponderador según elegibilidad mediana». Para los procedimientos agrupados, los ponderadores se normalizaron a una media de uno dentro de cada país, de modo que los países conservaran una influencia promedio equivalente sin perder la estructura de ponderación interna de cada muestra.

Esta decisión de ponderación exige justificación explícita. Los ponderadores originales reproducen poblaciones nacionales de establecimientos elegibles de magnitud muy dispar: México concentra el 84,9 % de la suma de ponderadores del conjunto agrupado, de suerte que su empleo directo convertiría el análisis regional en una estimación mexicana con contribuciones marginales de los otros cinco países. La normalización preserva íntegramente la corrección por probabilidad de selección y por no respuesta dentro de cada muestra, que es la función inferencial del ponderador, y neutraliza únicamente la escala poblacional relativa entre países. Las estimaciones nacionales de la Tabla 2 conservan por ello los ponderadores originales. Como verificación, los procedimientos agrupados se replicaron con los ponderadores sin normalizar: la correlación de rangos ponderada asciende a 0,417 —cifra que refleja el predominio del patrón mexicano, el más positivo de la muestra— y la partición resultante mantiene una concordancia elevada con la reportada (índice de Rand ajustado = 0,888).

Operacionalización de la preparación

La preparación digital se operacionalizó como capacidad operativa digital observable y no como una escala integral de transformación digital. Cuatro componentes armonizados se convirtieron a un rango común de 0 a 100: posesión de sitio web, conectividad y fiabilidad de internet, porcentaje de pagos de clientes recibidos por vía electrónica y porcentaje de pagos realizados por esa misma vía. La posesión de sitio web recibió 100 en caso afirmativo y 0 en caso negativo. El estado de la conexión a internet se puntuó con 0 cuando no existía conexión, con 50 cuando el establecimiento había experimentado interrupciones durante el ejercicio fiscal previo y con 100 cuando disponía de conexión sin interrupciones reportadas. Las variables de pago electrónico conservaron su escala porcentual original de 0 a 100. El Índice de Preparación Digital (IPD) corresponde a la media aritmética de los cuatro componentes en los casos completos.

El Índice de Preparación Digital se concibe explícitamente como un *índice formativo*, esto es, como un compuesto en el que los indicadores definen el constructo por agregación de facetas complementarias de la

capacidad operativa digital, y no como una escala reflectiva cuyos reactivos deban comportarse como manifestaciones intercambiables de un mismo factor latente (Diamantopoulos y Winklhofer, 2001). De esa naturaleza formativa se sigue que la consistencia interna elevada no constituye un criterio de calidad del índice, y que la eliminación o la reponderación de componentes poco correlacionados alteraría el dominio del constructo en lugar de depurarlo. La Preparación Operativa en Gestión de la Información, en cambio, se concibe como una escala reflectiva de rutinas directivas y se somete por ello a los criterios convencionales de unidimensionalidad y fiabilidad. Los contrastes factoriales reportados más adelante evalúan cada medida conforme al criterio que le corresponde.

La preparación operativa en gestión de la información se concibió como un constructo más estrecho, centrado en rutinas observables de seguimiento, gobernanza de metas y difusión interna de la información de desempeño. Se apoyó en cuatro componentes de prácticas de gestión del Banco Mundial derivados de las preguntas R.2 a R.7: número de indicadores de desempeño sometidos a seguimiento, horizonte temporal de las metas de producción o servicio, alcanzabilidad de las metas y conocimiento de las metas dentro de la organización. Las respuestas se mapearon a la escala 0–100 siguiendo las convenciones de codificación del Banco Mundial para prácticas estructuradas de gestión, en las que las prácticas más explícitas y estructuradas reciben puntuaciones superiores (Banco Mundial, 2026). El índice primario excluyó el componente R.1, relativo a la respuesta ante problemas, porque la opción «no aplica» puede señalar la ausencia de un problema operativo y convertirse así en una fuente de selección ajena a la capacidad de gestión de la información. El índice omite igualmente los reactivos vinculados a incentivos del módulo oficial de prácticas de gestión del Banco Mundial, de manera que cubre las familias de seguimiento y de fijación de metas, pero no la de incentivos (Viganola y Diallo, 2025). La puntuación de Preparación Operativa en Gestión de la Información (POGI) es la media de los cuatro componentes completos. Dado que el cuestionario aplica filtros de salto, los establecimientos que declaran no dar seguimiento a ningún indicador de desempeño (R.2) y no fijar metas de producción o servicio (R.4) no reciben los reactivos posteriores y obtienen cero en los cuatro componentes. La medida presenta, por tanto, un piso estructural: una puntuación de cero registra la ausencia declarada de rutinas formalizadas de seguimiento y fijación de metas, no un nivel bajo graduado de capacidad de gestión de la información.

Tabla 1.

Operacionalización de las dos dimensiones de preparación

Dimensión	Componente	Variable	Codificación
Preparación digital	Posesión de sitio web	C.22b	No = 0; Sí = 100
Preparación digital	Conectividad y fiabilidad de internet	C.39	Sin conexión = 0; con interrupciones = 50; sin interrupciones = 100
Preparación digital	Pagos recibidos por vía electrónica	K.33	Porcentaje observado, 0–100
Preparación digital	Pagos realizados por vía electrónica	K.38	Porcentaje observado, 0–100
Preparación en gestión de la información	Indicadores de desempeño con seguimiento	R.2–R.3	Codificación de prácticas estructuradas del Banco Mundial, 0–100
Preparación en gestión de la información	Horizonte temporal de las metas	R.4–R.5	Codificación de prácticas estructuradas del Banco Mundial, 0–100
Preparación en gestión de la información	Alcanzabilidad de las metas	R.4–R.6	Codificación de prácticas estructuradas del Banco Mundial, 0–100
Preparación en	Conocimiento de las	R.4–R.7	Codificación de prácticas estructuradas

Dimensión	Componente	Variable	Codificación
gestión de la información	metas		del Banco Mundial, 0–100

Fuente: operacionalización a partir de las variables de las Encuestas de Empresas del Banco Mundial 2023 y de las descripciones de indicadores del Banco Mundial.

Brecha de preparación y análisis estadístico

Para comparar la posición relativa de ambos dominios, los dos índices se estandarizaron sobre la muestra analítica agrupada con ponderadores normalizados por país. La Brecha de Preparación Digital-Informacional se definió como $BPD_i = z(IPD_i) - z(POGI_i)$, donde los valores positivos indican una posición digital relativa superior a la posición relativa del establecimiento en gestión de la información, y los negativos señalan lo contrario. La brecha constituye un contraste descriptivo entre dominios y no un índice universal validado de preparación ni una medida absoluta de capacidad de implementación.

Las medias nacionales de los índices y las estimaciones de la brecha se calcularon con ponderadores muestrales. Los intervalos de confianza se obtuvieron de 1000 remuestreos bootstrap ejecutados dentro de los estratos de muestreo de la Encuesta de Empresas de cada país, con lo que se preservó la estructura estratificada observada. La asociación monótona entre ambas dimensiones se examinó mediante la correlación de rangos de Spearman, con un intervalo de confianza bootstrap estratificado de modo análogo para el coeficiente agrupado. De forma adicional se calculó una correlación de rangos ponderada, definida como la correlación de Pearson ponderada de los rangos promedio con ponderadores normalizados por país.

Los estratos de remuestreo son los declarados por la encuesta, que combinan sector, tamaño y región dentro de cada país; la estratificación por país que se menciona en los procedimientos de conglomerados se refiere a la asignación del número de unidades remuestreadas entre países y no a un colapso de los estratos internos. Para despejar cualquier ambigüedad sobre este punto, el remuestreo agrupado se repitió con estratificación explícita por la interacción de país y sector, restringida a las celdas con al menos veinte observaciones ($n = 1987$): el intervalo resultante para la correlación de Spearman agrupada fue $[0,180; 0,265]$, prácticamente indistinguible del reportado en el texto principal.

El piso estructural de la medida informacional se abordó en dos planos complementarios. Primero, los análisis principales se replicaron sobre la submuestra que excluye los establecimientos con puntuación nula en los cuatro componentes, a fin de establecer si la asociación entre dominios y la estructura de conglomerados subsisten cuando se retira la masa de probabilidad situada en el origen. Segundo, la medida se descompuso conforme a una especificación en dos partes, en la tradición de los modelos de decisión secuencial para variables dependientes limitadas (Cragg, 1971): un componente binario que modela la probabilidad de declarar alguna rutina de seguimiento o de fijación de metas, estimado por regresión logística ponderada, y un componente de intensidad que modela por mínimos cuadrados ponderados la puntuación condicional entre los establecimientos con rutina declarada. Ambos componentes incorporan el mismo vector de características organizativas, el país y el sector, lo que permite establecer si los determinantes de la presencia de gestión informacional formalizada coinciden con los de su grado de desarrollo.

La estructura dimensional de cada conjunto de componentes se examinó mediante un modelo confirmatorio de un factor. Puesto que los cuatro componentes informacionales son ordinales con cuatro o cinco categorías, la matriz de entrada fue la de correlaciones policóricas estimadas en dos pasos, procedimiento preferible al tratamiento continuo en presencia de indicadores ordinales (Li, 2016); el modelo se ajustó por mínimos cuadrados y se reportan las cargas estandarizadas, la varianza media extraída, la fiabilidad compuesta, la raíz del residuo cuadrático medio estandarizado y los autovalores de la matriz de entrada. El mismo procedimiento se aplicó a los cuatro componentes digitales, con una versión sobre correlaciones de Pearson y otra sobre correlaciones policóricas tras discretizar los porcentajes de pago en cuatro categorías ordenadas. Para el Índice de Preparación Digital se estimó además una variante ponderada por las cargas factoriales, contrastada con la media de pesos iguales mediante la correlación entre ambas versiones y la concordancia de las particiones que cada una induce.

Conglomerados y validación de perfiles

Los perfiles de preparación se estimaron a partir de los valores estandarizados del Índice de Preparación Digital y de la Preparación Operativa en Gestión de la Información mediante K-medias ponderadas. La estandarización utilizó ponderadores muestrales normalizados por país, y el algoritmo se ajustó con esos mismos ponderadores, cien inicializaciones y una semilla aleatoria fija que garantiza la reproducibilidad. Las soluciones candidatas de dos a seis conglomerados se compararon mediante el coeficiente de silueta, el índice de Calinski-Harabasz, el índice de Davies-Bouldin y la inercia intraconglomerado. La estabilidad se evaluó con 200 remuestreos bootstrap estratificados por país: cada solución bootstrap se ajustó a los datos remuestreados y se proyectó sobre la muestra original, y la concordancia con la partición original se cuantificó con el índice de Rand ajustado (Hubert y Arabie, 1985). El coeficiente de silueta operó como criterio primario de selección porque captura simultáneamente la cohesión interna y la separación entre conglomerados (Rousseeuw, 1987). Estos índices comparan particiones candidatas entre sí y cuantifican la reproducibilidad de una partición dada bajo remuestreo, pero no contrastan el número de conglomerados retenido frente a un modelo nulo de distribución continua única. Los perfiles obtenidos se tratan, por consiguiente, como una partición descriptiva reproducible del espacio conjunto de preparación y no como evidencia de subpoblaciones naturalmente discretas.

Para someter a prueba el supuesto de conglomerados esféricos y de tamaño comparable que subyace a K-medias, la partición retenida se comparó con dos alternativas. La primera fue un modelo de mezclas gaussianas con covarianzas específicas por componente y selección del número de componentes por el criterio de información bayesiano (Fraley y Raftery, 2002). La segunda fue un algoritmo de conglomerados basado en densidad, que no fija el número de grupos a priori y aísla las observaciones no asignables como ruido, explorado sobre una rejilla de radios de vecindad (Ester et al., 1996). La concordancia entre soluciones se cuantificó de nuevo con el índice de Rand ajustado.

Las etiquetas de los conglomerados se asignaron solo después de la estimación, a partir de sus centroides ponderados en las escalas originales de 0 a 100. Este procedimiento evitó imponer cuadrantes a priori y permitió que la estructura empírica determinara si las configuraciones teóricamente plausibles estaban realmente presentes.

Análisis mediante aprendizaje automático explicable

Se recurrió a modelos supervisados para determinar si los perfiles derivados empíricamente podían distinguirse a partir de características organizativas no empleadas en la construcción de ninguno de los dos índices. Los predictores candidatos fueron el tamaño y la antigüedad de la empresa, la experiencia del directivo principal, su sexo, la propiedad extranjera, la actividad exportadora, la certificación internacional de calidad, la capacitación formal de los trabajadores, la titularidad de cuenta bancaria, el acceso al crédito, el sector económico y el país. La exclusión de los componentes de preparación del conjunto de predictores impidió la filtración directa de información.

Se evaluaron una regresión logística multinomial, bosques aleatorios y potenciación extrema del gradiente (XGBoost) mediante validación cruzada estratificada de cinco pliegues. Para evitar la filtración derivada de la construcción de los perfiles, el análisis de conglomerados se reajustó en cada pliegue de entrenamiento y los establecimientos de prueba se asignaron a los centroides más próximos obtenidos del entrenamiento antes de puntuar los clasificadores. Las variables numéricas se imputaron por la mediana y se estandarizaron cuando correspondía; las binarias y categóricas se imputaron por la moda, y los factores categóricos recibieron codificación one-hot. La valoración de los modelos privilegió la exactitud balanceada, el F1 macropromediado y la pérdida logarítmica multiclase por encima de la exactitud global, dado el tamaño desigual de los perfiles de preparación. Los bosques aleatorios emplearon 500 árboles con submuestreo balanceado por clase, mientras que XGBoost utilizó 400 árboles con ajustes conservadores de profundidad y tasa de aprendizaje (Breiman, 2001; Chen y Guestrin, 2016).

El desempeño de clasificación se contrastó frente a un modelo nulo de ausencia de estructura mediante pruebas de permutación (Ojala y Garriga, 2010). El esquema empleado es una aleatorización restringida: en cada réplica se permutan conjuntamente las filas del par de índices de preparación y de su ponderador, de modo que la geometría del espacio conjunto y la estructura de conglomerados se conservan intactas mientras se rompe el vínculo entre la posición de preparación y las características organizativas del establecimiento. Sobre cada muestra permutada se reejecuta íntegramente el procedimiento anidado —reajuste de

conglomerados dentro de cada pliegue de entrenamiento, asignación de los casos de prueba y estimación de los tres clasificadores— y se registra la exactitud balanceada. El valor de p empírico es la proporción de réplicas nulas que igualan o superan el estadístico observado. El mismo esquema se aplicó a la validación con exclusión sucesiva de países, donde la proximidad de los valores observados al umbral de azar de un tercio para tres clases hace indispensable el contraste formal; en ese caso las permutaciones se ejecutaron dentro de cada país, con lo que se preserva la composición nacional de perfiles, y los valores de p por país se combinaron mediante el procedimiento de Fisher. Para mantener la carga computacional dentro de límites razonables, las réplicas nulas y el estadístico observado con el que se comparan se calcularon con una configuración aligerada de los modelos de árboles (120 árboles), cuya exactitud balanceada reproduce la de la configuración plena con diferencias inferiores a 0,005.

La interpretación de los modelos se apoyó en las explicaciones aditivas de Shapley (SHAP), que asignan contribuciones de predicción a variables individuales y admiten agregación para caracterizar el comportamiento global del modelo (Lundberg y Lee, 2017; Lundberg et al., 2020). A lo largo de este artículo, la expresión aprendizaje automático explicable designa esta atribución de Shapley ex post aplicada a un clasificador de conjuntos de árboles, y no una clase más amplia de modelos interpretables por diseño. Puesto que los bosques aleatorios admiten una estimación eficiente de valores de Shapley específica para árboles sin renunciar a las interacciones no lineales, la capa explicativa se construyó con un modelo de bosque aleatorio ajustado sobre la totalidad de los datos, una vez evaluado el desempeño predictivo mediante construcción anidada de los perfiles y validación cruzada. Los valores absolutos medios de Shapley se resumieron por familia de predictores para cada perfil, sumando las contribuciones de los indicadores one-hot pertenecientes a una misma variable original. Como esa agregación suma a través de categorías, las familias con más niveles —país, con seis niveles, y sector, con ocho— acumulan totales mayores que los predictores binarios; las magnitudes agrupadas no resultan directamente comparables entre predictores de distinta cardinalidad y se reportan como ordenamiento interno, no entre tipos de predictor. Para corregir esa falta de comparabilidad, cada magnitud agrupada se reporta también dividida por el número de niveles de la variable de origen, lo que devuelve una contribución media por indicador y permite el contraste directo entre los factores politómicos y los predictores binarios. Estos valores se interpretaron como medidas de atribución del modelo y no como efectos causales. La capacidad de generalización entre contextos nacionales se evaluó mediante validación con exclusión sucesiva de países: para cada país excluido se reajustaron tanto el análisis de conglomerados como el clasificador sobre los cinco países restantes, y se omitió del conjunto de predictores el propio indicador de país.

La colinealidad entre los predictores organizativos se diagnosticó sobre la matriz de diseño de la regresión logística multinomial. Puesto que el país y el sector son factores politómicos cuyos indicadores no admiten lectura individual del factor de inflación de la varianza, se calculó el factor generalizado de inflación de la varianza, que agrega los indicadores pertenecientes a un mismo factor y resulta invariante a la codificación elegida (Fox y Monette, 1992). Se reporta la versión corregida por grados de libertad, comparable con el umbral convencional de 2,24 que corresponde a un factor de inflación de cinco en predictores de un solo grado de libertad.

Análisis de sensibilidad y gobernanza de datos

Los análisis de robustez examinaron una especificación alternativa de gestión de la información que incorporó el componente R.1 de respuesta ante problemas, una puntuación más estricta de internet que asignó cero a las empresas con interrupciones reportadas, una solución no ponderada de tres conglomerados y análisis de componentes principales separados para los componentes del Índice de Preparación Digital y de la Preparación Operativa en Gestión de la Información. La validez convergente del indicador indirecto de gestión de la información se contrastó además con una versión reconstruida del índice completo de ocho componentes de prácticas de gestión del Banco Mundial, entre los establecimientos que disponían de todos los reactivos requeridos. Se añadió el examen de la heterogeneidad sectorial dentro de El Salvador, cuya brecha estandarizada resultó la más negativa de la muestra, mediante estimaciones ponderadas de ambos índices y de la brecha por sector a un dígito y una prueba de Kruskal-Wallis sobre la brecha, complementadas con la comparación del nivel digital salvadoreño frente al de los mismos sectores en los otros cinco países.

El análisis utilizó microdatos secundarios anonimizados obtenidos conforme a las condiciones de acceso de las Encuestas de Empresas del Banco Mundial. No se accedió a información personal identificable, no se

reclutó ni contactó a participantes y los resultados se presentan únicamente de forma agregada. Los microdatos originales continúan sujetos al protocolo de acceso del Banco Mundial; los archivos de análisis derivados, las salidas estadísticas, las figuras y el código de reproducibilidad pueden acompañar al manuscrito como material complementario de investigación. Todos los análisis se realizaron en Python con bibliotecas estándar de computación científica y aprendizaje automático. Las pruebas estadísticas fueron bilaterales y se consideró significativo un valor de $p < 0,05$.

RESULTADOS

Alineación entre preparación digital y preparación en gestión de la información

En los 2143 establecimientos analizados, la preparación digital y la preparación operativa en gestión de la información se asociaron positivamente, si bien con intensidad débil (ρ de Spearman sin ponderar = 0,213; IC 95 % por bootstrap estratificado [0,175; 0,244]; $p < 0,001$). La correlación de rangos ponderada y normalizada por país alcanzó 0,271, lo que indica que la ponderación muestral refuerza ligeramente la asociación sin modificar su clase de magnitud. Tal magnitud revela que ambos constructos comparten cierto ordenamiento organizativo, pero no admiten tratamiento como indicadores equivalentes de modernización empresarial. Las asociaciones nacionales fueron heterogéneas: México presentó la relación ponderada de rangos más intensa y positiva ($r = 0,470$), en tanto que Paraguay mostró una relación negativa ($r = -0,352$). Atendiendo a la incertidumbre que introducen estas submuestras, los coeficientes nacionales se acompañan ahora de intervalos bootstrap estratificados de 2000 réplicas (Tabla 2). El coeficiente paraguayo, el único negativo, resultó estadísticamente distinguible de cero (IC 95 % [-0,484; -0,184]), mientras que el costarricense fue indistinguible de cero ($r = -0,011$; IC 95 % [-0,187; 0,164]). El ordenamiento fino entre países sigue mereciendo lectura descriptiva, pero el contraste entre el patrón mexicano y el paraguayo no es atribuible al error de muestreo. Aun así, los contrastes resultan coherentes con el tratamiento de la capacidad operativa digital y de la gestión basada en información como dimensiones distintas.

Tabla 2.

Correlaciones de rangos ponderadas entre dominios, por país, con intervalos bootstrap

País	n	r de rangos ponderada	IC 95 % bootstrap estratificado	% con POGI nula
México	682	0,470	[0,347; 0,603]	18,5
Perú	364	0,312	[0,163; 0,495]	5,8
Colombia	443	0,161	[0,024; 0,293]	7,7
El Salvador	285	0,131	[-0,033; 0,282]	23,2
Costa Rica	177	-0,011	[-0,187; 0,164]	14,7
Paraguay	192	-0,352	[-0,484; -0,184]	14,6

Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023. Nota: 2000 remuestreos dentro de los estratos de muestreo de cada país. La última columna indica la proporción de establecimientos con puntuación nula en los cuatro componentes de gestión de la información.

Sensibilidad a los ceros estructurales. La exclusión de los 301 establecimientos que puntúan cero en los cuatro componentes informacionales —el 14,05 % de la muestra— redujo el conjunto a 1842 casos y atenuó la asociación entre dominios sin alterar su signo ni su significación: ρ de Spearman = 0,168 (IC 95 % [0,124; 0,208]; $p < 0,001$) y correlación de rangos ponderada de 0,174. La bimodalidad inducida por los filtros de salto del cuestionario contribuye, por tanto, a la asociación observada, pero no la produce.

Descomposición en dos partes. La especificación secuencial separa con nitidez dos procesos. En el componente de presencia, la probabilidad de declarar alguna rutina formalizada de seguimiento o de fijación de metas se asoció sobre todo con la capacitación formal de los trabajadores (razón de momios = 5,24), la

titularidad de cuenta bancaria (3,45), la propiedad extranjera (2,39), la actividad exportadora (2,18), el acceso al crédito (1,95), la certificación de calidad (1,64) y el tamaño (1,55 por unidad logarítmica de empleo). En el componente de intensidad, estimado sobre los 1842 establecimientos con rutina declarada ($R^2 = 0,236$), el ordenamiento cambia: la titularidad de cuenta bancaria ($\beta = 18,85$ puntos) y la capacitación formal ($\beta = 10,03$) dominan, seguidas de la propiedad extranjera (5,55), la certificación (5,26) y el tamaño (3,72); el sexo femenino del directivo principal se asoció de forma negativa ($\beta = -4,69$), asociación ausente en el componente de presencia. Los dos procesos comparten así buena parte de sus determinantes, pero no su jerarquía, lo que desaconseja tratar la puntuación de gestión informacional como un continuo homogéneo.

Estructura factorial de las dos medidas

El modelo confirmatorio de un factor sobre la matriz policórica de los cuatro componentes informacionales arrojó cargas estandarizadas de 0,539 para el seguimiento de indicadores, 0,874 para el horizonte de las metas, 0,723 para su alcanzabilidad y 0,761 para su conocimiento interno, con una varianza media extraída de 0,539, una fiabilidad compuesta de 0,820, una raíz del residuo cuadrático medio estandarizado de 0,030 y un residuo máximo de 0,044. El primer autovalor concentra el 64,4 % de la varianza y cuadruplica al segundo (2,576 frente a 0,658). La Preparación Operativa en Gestión de la Información se comporta, en consecuencia, como una escala reflectiva unidimensional con fiabilidad aceptable, lo que respalda su cálculo como media de los cuatro componentes.

El mismo modelo aplicado a los cuatro componentes digitales produce un resultado opuesto y obliga a precisar la naturaleza del índice. Las correlaciones entre componentes son casi nulas salvo entre los dos reactivos de pago electrónico: 0,032 entre sitio web y conectividad, $-0,009$ y $-0,007$ entre conectividad y los pagos recibidos y realizados, y 0,520 entre ambos flujos de pago. La solución de un factor asigna cargas de 0,084 al sitio web, $-0,007$ a la conectividad, 0,520 a los pagos recibidos y 0,999 a los realizados, con una varianza media extraída de 0,319, una fiabilidad compuesta de 0,483 y dos autovalores superiores a la unidad (1,535 y 1,028); el primer factor explica solo el 38,4 % de la varianza y el ajuste es deficiente bajo criterios reflectivos (CFI = 0,845; TLI = 0,536; RMSEA = 0,154). La versión policórica, obtenida tras discretizar los porcentajes de pago en cuatro categorías ordenadas, reproduce el patrón (varianza media extraída de 0,336; primer factor con el 40,3 %). Parte de esta ausencia de covariación es atribuible a la escasa variabilidad de dos componentes en esta población: el 87,3 % de los establecimientos posee sitio web y solo el 0,4 % carece de conexión a internet, de modo que la conectividad discrimina casi exclusivamente entre servicio interrumpido y no interrumpido.

La variante del Índice de Preparación Digital ponderada por las cargas factoriales ilustra el precio de forzar una lectura reflectiva: concentra el 94,7 % del peso en los dos reactivos de pago (0,324 y 0,623), asigna peso nulo a la conectividad, correlaciona solo 0,785 con la media de pesos iguales y genera una partición apreciablemente distinta de la reportada (índice de Rand ajustado = 0,700). Antes que una depuración del índice, la reponderación equivaldría a redefinirlo como una medida de digitalización de los flujos de pago. Se conserva por ello la media de pesos iguales, y se declara de manera explícita que el Índice de Preparación Digital es un compuesto formativo cuyos componentes cubren facetas complementarias y no intercambiables de la capacidad operativa digital (Diamantopoulos y Winklhofer, 2001). Las consecuencias interpretativas de esta precisión se retoman en la discusión y en las limitaciones.

Estimaciones nacionales de preparación

Las estimaciones nacionales ponderadas mostraron puntuaciones digitales comparativamente altas en Costa Rica, Colombia y Perú, mientras las puntuaciones de gestión de la información fueron sistemáticamente menores y más dispersas. Perú registró la preparación ponderada más alta en gestión de la información (57,86), seguido de Costa Rica (54,14) y Colombia (53,52). Tras la estandarización, Colombia presentó la mayor brecha nacional positiva (0,173), en contraste con la marcada brecha negativa de El Salvador ($-0,559$). Los intervalos de confianza señalaron que la dirección de la brecha se separa de cero con mayor nitidez en Colombia y El Salvador.

Tabla 3.*Estimaciones ponderadas de preparación por país*

País	n	Preparación digital, IC 95 %	Preparación en gestión de la información, IC 95 %	Brecha estandarizada, IC 95 %
Colombia	443	84,43 [82,35; 86,42]	53,52 [50,14; 56,63]	0,173 [0,026; 0,306]
Costa Rica	177	84,67 [82,43; 86,91]	54,14 [48,41; 60,38]	0,169 [-0,103; 0,424]
El Salvador	285	65,75 [63,08; 68,47]	38,32 [33,79; 42,96]	-0,559 [-0,738; -0,382]
México	682	79,52 [78,04; 81,03]	46,95 [39,97; 53,63]	0,065 [-0,123; 0,264]
Paraguay	192	78,63 [75,72; 81,11]	44,73 [39,58; 49,51]	0,078 [-0,225; 0,388]
Perú	364	83,67 [81,36; 86,17]	57,86 [55,00; 60,83]	-0,018 [-0,204; 0,173]

Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023. Nota: estimaciones con los ponderadores muestrales originales; decimales alineados.

Heterogeneidad sectorial en El Salvador. La brecha salvadoreña, la más negativa de la muestra, no es homogénea entre ramas de actividad (Kruskal-Wallis $H = 13,49$; $p = 0,004$ sobre cuatro sectores con al menos diez observaciones). La manufactura, que concentra 153 de los 285 establecimientos, presenta la brecha más negativa (-1,024) al combinar la puntuación informacional más alta del país (48,99) con una puntuación digital de 63,97. La construcción muestra el patrón inverso y más severo en términos absolutos: una puntuación informacional de 19,55, con el 58,3 % de sus establecimientos en el piso estructural. El comercio mayorista y minorista se aproxima al equilibrio (-0,068). Un contraste adicional despeja la interpretación de la brecha agregada: El Salvador registra la puntuación digital ponderada más baja de los seis países en los cuatro sectores examinados. La brecha negativa proviene, por tanto, de un déficit digital generalizado y no de una capacidad informacional comparativamente elevada, salvo en la manufactura, donde ambos elementos concurren.

Perfiles empíricos de preparación

La solución de tres conglomerados produjo la partición global más sólida, con un coeficiente de silueta de 0,463, un índice de Calinski-Harabasz de 1834,80 y un índice de Davies-Bouldin de 0,796. El índice de Rand ajustado promedio bajo bootstrap fue de 0,909, con un percentil cinco de 0,799, lo que denota una elevada estabilidad ante el remuestreo. La solución de dos conglomerados exhibió menor separación y una concordancia bootstrap media más baja, mientras que las soluciones de cuatro a seis conglomerados arrojaron valores de silueta inferiores y una estabilidad sensiblemente peor.

El perfil más numeroso reunió a 1300 líderes de preparación dual, con centroides ponderados de 86,74 en preparación digital y 69,02 en gestión de la información. Un segundo grupo de 429 establecimientos resultó digitalmente avanzado pero restringido en el manejo de la información, con un centroe digital de 81,13 frente a uno de apenas 10,09 en gestión de la información. Los 414 establecimientos restantes conformaron el grupo de rezagados en ambas dimensiones, con centroides de 54,32 y 39,26. El bajo centroe del segundo perfil refleja el piso estructural de la medida: 301 establecimientos, equivalentes al 14,0 % de la muestra, obtuvieron cero en los cuatro componentes de gestión de la información, y 236 de ellos —el 55,0 % de ese perfil— quedaron clasificados en el grupo digitalmente avanzado con restricción informacional. El perfil identifica así a empresas que declaran no aplicar rutina alguna de seguimiento o fijación de metas, y no a empresas distribuidas a lo largo de un rango bajo de la escala. En la solución de tres conglomerados seleccionada, los datos no generaron un conglomerado diferenciado de empresas con capacidad informacional y restricción digital.

Reestimación sin ceros estructurales. La pregunta de si este segundo perfil constituye una configuración organizativa o un subproducto de los filtros de salto admite respuesta directa. Reestimados los conglomerados sobre los 1842 establecimientos con al menos una rutina declarada, la solución de tres grupos sigue siendo la preferida por el coeficiente de silueta (0,406 frente a 0,395 para dos grupos), y los tres perfiles conservan su identidad cualitativa: 995 líderes duales (87,26 y 76,48), 469 establecimientos digitalmente avanzados con restricción informativa (85,27 y 38,67) y 378 establecimientos con restricción digital (57,01 y 55,81). El perfil digitalmente avanzado persiste, pero con un centroide informativo casi cuatro veces superior al de la solución principal, y la silueta desciende de 0,463 a 0,406. La concordancia entre ambas particiones sobre los casos comunes es moderada (índice de Rand ajustado = 0,549). Dos conclusiones se siguen de aquí. La configuración de alta capacidad digital con baja formalización directiva es sustantiva y no un artefacto del instrumento; pero la nitidez de la separación reportada en la solución principal está considerablemente inflada por la concentración de casos en el origen, y el centroide de 10,09 debe leerse como una propiedad de la medida antes que como una descripción del nivel informativo típico de esas empresas. Conviene notar, además, que el tercer conglomerado de esta reestimación —digitalmente restringido y con capacidad informativa cercana al promedio— se aproxima a la configuración cuya ausencia se comenta más adelante.

Tabla 4.

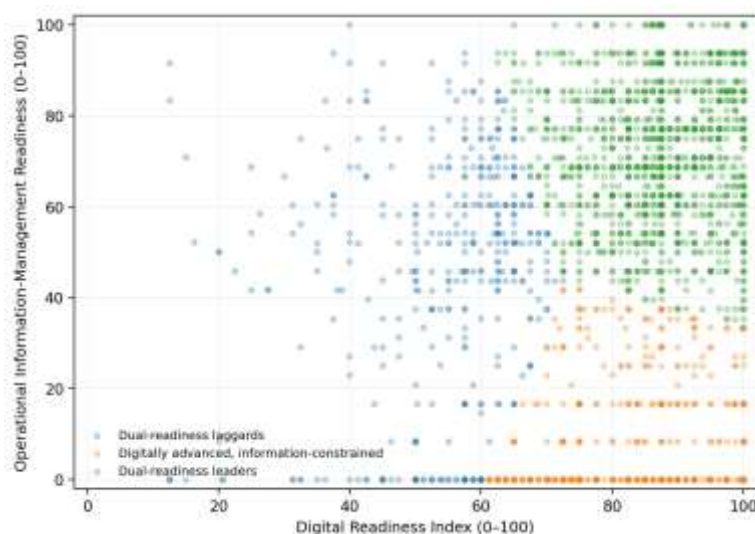
Perfiles empresariales de preparación derivados empíricamente

Perfil	n	Part. ponderada (%)	Preparación digital	Gestión de la información	Brecha estand.
Líderes de preparación dual	1300	58,2	86,74	69,02	−0,180
Digitalmente avanzados con restricción informativa	429	24,5	81,13	10,09	1,372
Rezagados en ambas dimensiones	414	17,3	54,32	39,26	−1,341
Sin ceros estructurales: líderes duales	995	49,9	87,26	76,48	—
Sin ceros estructurales: digitalmente avanzados con restricción informativa	469	33,2	85,27	38,67	—
Sin ceros estructurales: restricción digital	378	16,9	57,01	55,81	—

Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023. Nota: la banda horizontal de observaciones situada en cero sobre el eje vertical corresponde a los establecimientos que no reportan indicadores con seguimiento ni metas, y cuyos cuatro componentes de gestión de la información reciben por ello puntuación cero. La brecha estandarizada es una función determinista de las dos variables empleadas para estimar los conglomerados y se reporta únicamente para facilitar la lectura descriptiva. Las tres últimas filas corresponden a la reestimación sobre los 1842 establecimientos con al menos una rutina declarada.

Figura 1.

Distribución conjunta de la preparación digital y de la preparación en gestión de la información



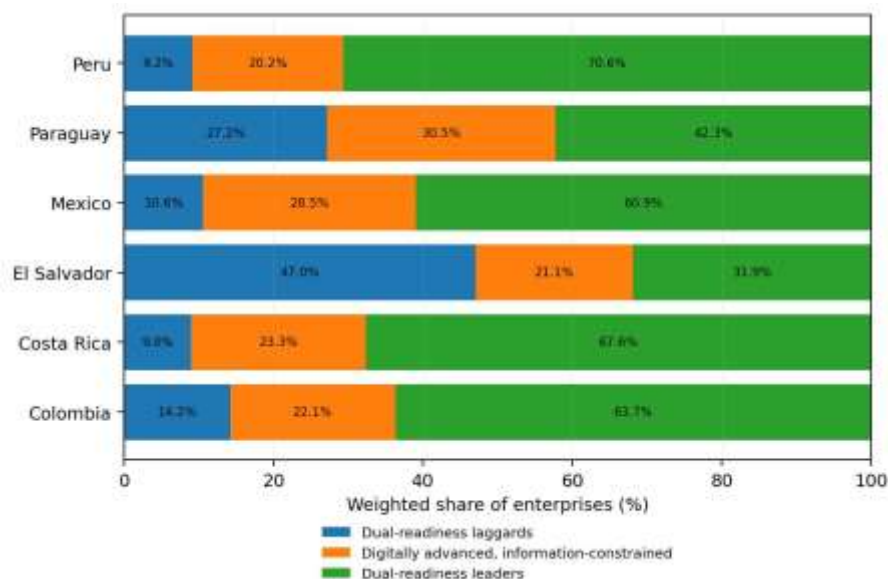
Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023

Especificaciones alternativas de conglomerado. El contraste con dos familias distintas de algoritmos delimita el estatus de la partición. El modelo de mezclas gaussianas no respalda la solución de tres componentes: el criterio de información bayesiano desciende de manera monótona en el rango explorado, y la concordancia con K-medias es de 0,551 en $k = 3$ y de 0,719 en $k = 2$. El algoritmo basado en densidad resulta más elocuente todavía: con radios de vecindad de 0,20 y 0,25 identifica tres regiones densas, con 497 y 263 observaciones no asignables respectivamente y concordancias de 0,573 y 0,557, pero a partir de un radio de 0,30 la nube se resuelve en un único conglomerado conexo. Los datos no contienen, por consiguiente, subpoblaciones separadas por regiones de baja densidad. Este resultado no invalida la partición reportada — que sigue siendo la descripción reproducible más informativa del espacio conjunto— pero convierte la advertencia formulada en la metodología en un hallazgo respaldado: los perfiles son cortes útiles de un continuo, no tipos organizativos discretos.

La composición nacional difirió de manera apreciable entre perfiles. Perú concentró la mayor participación ponderada de líderes de preparación dual (70,6 %), seguido de Costa Rica (67,6 %), Colombia (63,7 %) y México (60,9 %). Paraguay registró la proporción más alta de empresas digitalmente avanzadas con restricción informacional (30,5 %), seguido de México (28,5 %). El Salvador exhibió la composición más adversa, con un 47,0 % de establecimientos clasificados como rezagados en ambas dimensiones y un 31,9 % como líderes.

Figura 2.

Distribución ponderada de los perfiles de preparación por país



Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023. Nota: la nota metodológica sobre agregación de valores de Shapley que acompañaba a esta figura correspondía en realidad a la Figura 3 y se ha trasladado a su lugar. Figura provista a 300 ppp.

Discriminación predictiva y explicabilidad

El desempeño de clasificación resultó moderado antes que elevado una vez reajustados los perfiles de preparación dentro de cada pliegue de entrenamiento. La regresión logística multinomial alcanzó la mayor exactitud balanceada media (0,575) y el F1 macropromediado más alto (0,548). Los bosques aleatorios se mantuvieron competitivos (exactitud balanceada de 0,546; F1 macro de 0,539), mientras XGBoost produjo la menor pérdida logarítmica (0,808) con una discriminación más débil según las métricas balanceadas por clase. Estos resultados indican que los perfiles de preparación solo son parcialmente recuperables a partir de características generales de la empresa, lo que respalda la interpretación de que los índices captan estados organizativos irreducibles al tamaño, el sector o la orientación internacional.

Contraste frente al modelo nulo. La moderación de estas cifras planteaba la pregunta de si superan lo esperable por azar. Las pruebas de permutación restringida, con 507 réplicas en las que se rompe el vínculo entre la posición de preparación y las características organizativas preservando la geometría del espacio conjunto, responden de forma inequívoca. La distribución nula de la exactitud balanceada se centra en 0,334, valor que coincide con el umbral teórico de un tercio para tres clases, con desviaciones típicas entre 0,003 y 0,015 y máximos de 0,381, 0,366 y 0,344 para la regresión logística, los bosques aleatorios y XGBoost. Los tres estadísticos observados —0,574, 0,550 y 0,491— superan el máximo de sus respectivas distribuciones nulas, de modo que el valor de p empírico alcanza el mínimo resoluble con ese número de réplicas ($p = 0,002$). El poder discriminante es, por tanto, modesto en magnitud absoluta pero incuestionable en significación.

Tabla 5.*Validación cruzada de cinco pliegues de los clasificadores de perfiles de preparación*

Modelo	Exactitud balanceada	F1 macro	Pérdida log.	Media nula y p (permutación)
Regresión logística	0,575 ± 0,049	0,548 ± 0,047	0,922 ± 0,048	0,334; p = 0,002
Bosques aleatorios	0,546 ± 0,044	0,539 ± 0,042	0,855 ± 0,041	0,333; p = 0,002
XGBoost	0,487 ± 0,039	0,499 ± 0,045	0,808 ± 0,043	0,334; p = 0,002

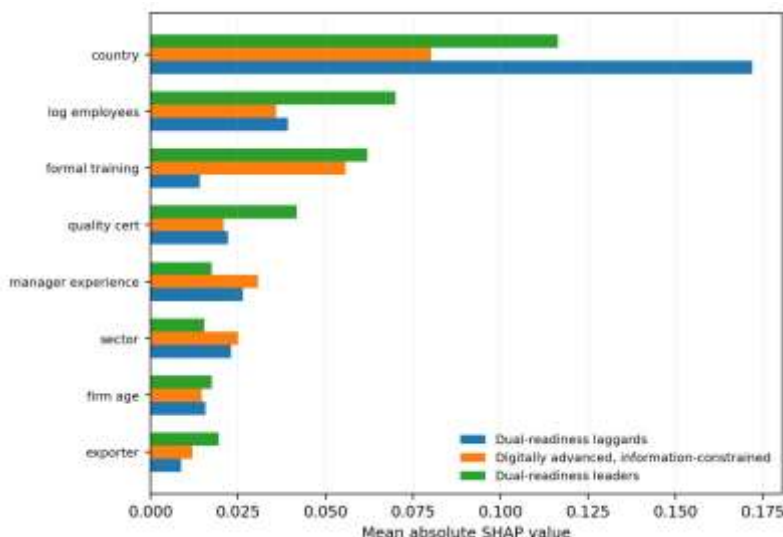
Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023. Nota: la distribución nula procede de 507 permutaciones restringidas del par de índices de preparación; p = 0,002 es el mínimo resoluble con ese número de réplicas. Decimales alineados.

Los contrastes descriptivos y la atribución de Shapley convergieron en varios factores organizativos. Los líderes de preparación dual presentaron frecuencias más altas de capacitación formal (67,1 %), certificación internacional de calidad (27,7 %), actividad exportadora (24,6 %) y propiedad extranjera (17,5 %) que los demás perfiles. En la explicación del bosque aleatorio, el contexto nacional aportó la mayor contribución agrupada de SHAP absoluto medio en las tres clases; dado que los valores agrupados suman las categorías one-hot, parte de esa magnitud responde a la cardinalidad de seis niveles de la variable país, y los resultados de exclusión sucesiva de países que se reportan más adelante proporcionan la evidencia independiente sobre el peso del contexto nacional. El tamaño de la empresa, la capacitación formal, la certificación de calidad, la experiencia directiva y el sector también destacaron, con una influencia particularmente marcada de la capacitación formal para separar a las empresas digitalmente avanzadas con restricción informacional de las líderes. La dirección y la magnitud de estas atribuciones deben leerse como estructura predictiva y no como influencia causal.

Atribuciones normalizadas por cardinalidad. La corrección por número de niveles reordena de manera sustancial esa jerarquía y desactiva la principal objeción interpretativa. Dividida la magnitud agrupada entre los niveles de cada familia, el país desciende del primer puesto al octavo de doce en el perfil digitalmente avanzado con restricción informacional y al séptimo entre los líderes duales, y solo conserva una posición alta —la segunda— entre los rezagados. La contribución media por indicador queda encabezada por la capacitación formal en el perfil digitalmente avanzado (0,055 frente a 0,012 del país) y por el tamaño de la empresa en los otros dos perfiles (0,080 entre los líderes y 0,044 entre los rezagados), con la certificación de calidad y la experiencia directiva en posiciones intermedias. El sector, con ocho niveles, abandona el primer cuartil en las tres clases. Buena parte de la preeminencia del contexto nacional en las magnitudes agrupadas era, por tanto, un efecto aritmético de la codificación; el peso genuino del país queda documentado por la validación con exclusión sucesiva y no por la atribución de Shapley. La Figura 3 incorpora un segundo panel con estos valores normalizados.

Figura 3.

Importancia global de las variables según valores de Shapley, por perfil de preparación (panel A: magnitudes agrupadas; panel B: contribución media por indicador)



Fuente: elaboración de los autores a partir de los microdatos de las Encuestas de Empresas del Banco Mundial 2023. Nota: los valores agrupados del panel A corresponden a la suma de los valores absolutos medios de Shapley de los indicadores one-hot de cada variable; país (seis niveles) y sector (ocho niveles) acumulan por ello totales mayores que los predictores binarios. El panel B divide cada total entre el número de niveles de la variable de origen. Figura provista a 300 ppp.

La validación con exclusión sucesiva de países impuso una restricción adicional a la generalización regional. La exactitud balanceada media fue de 0,468 para la regresión logística multinomial, de 0,447 para los bosques aleatorios y de 0,391 para XGBoost cuando cada país no observado se predijo a partir de modelos y centroides de preparación derivados de los cinco países restantes. Existen, por tanto, regularidades a escala de empresa, pero el contexto nacional sigue siendo un componente relevante de las configuraciones de preparación observadas.

Puesto que estos valores se aproximan al umbral de azar, se sometieron al mismo contraste de permutación, ejecutado dentro de cada país para preservar su composición de perfiles (517 réplicas). La media nula se situó de nuevo en torno a 0,334. Con la regresión logística, los seis países superaron significativamente el azar, con valores de p entre 0,002 y 0,010 y una combinación de Fisher de $\chi^2(12) = 71,78$ ($p = 1,5 \times 10^{-10}$); los bosques aleatorios arrojaron el mismo patrón y una combinación idéntica en magnitud. XGBoost resultó significativo en cinco de los seis países ($\chi^2(12) = 58,88$; $p = 3,6 \times 10^{-8}$), con la excepción de Paraguay ($p = 0,064$). El deterioro respecto de la validación cruzada interna es real y considerable, pero la señal transferible entre contextos nacionales no es nula: los modelos entrenados en cinco países predicen el sexto por encima del azar de forma sistemática, aunque con un margen que desaconseja cualquier uso predictivo aplicado.

Colinealidad. El diagnóstico descartó que la coexistencia de país, sector y tamaño en la matriz de diseño distorsione las estimaciones. Los factores generalizados de inflación de la varianza, corregidos por grados de libertad, se mantuvieron entre 1,016 y 1,196, con el máximo en la experiencia del directivo principal; el país, con cinco grados de libertad, alcanzó 1,070 y el sector, con siete, 1,029. Todos los valores quedan muy por debajo del umbral convencional de 2,24.

Los hallazgos principales se mantuvieron estables en las verificaciones de sensibilidad. La incorporación del componente de respuesta ante problemas arrojó una correlación de Spearman de 0,200 en 1928 establecimientos completos, mientras que una definición más estricta de internet redujo el coeficiente a 0,141 sin alterar su dirección positiva ni su significación estadística. La puntuación de Preparación Operativa en

Gestión de la Información mostró una fuerte asociación convergente con el índice completo de ocho componentes de prácticas de gestión del Banco Mundial reconstruido para este fin ($\rho = 0,866$; $n = 1659$; $p < 0,001$). La solución no ponderada de tres conglomerados resultó además altamente concordante con la partición ponderada (índice de Rand ajustado = 0,884). Los análisis de componentes principales por separado revelaron cargas positivas equilibradas para los cuatro componentes de gestión de la información, en tanto que el primer componente digital estuvo dominado por las variables de pago electrónico; ese resultado justificó conservar el índice digital transparente de pesos iguales en lugar de sustituirlo por un componente empírico que subrepresentaría los indicadores de sitio web y conectividad. El análisis factorial confirmatorio reportado más arriba precisa el alcance de esa observación: el predominio de los reactivos de pago no es una asimetría de grado sino la ausencia de un factor común a los cuatro componentes, lo que obliga a tratar el índice como compuesto formativo.

DISCUSIÓN

El hallazgo principal es que la capacidad operativa digital y la preparación operativa en gestión de la información están relacionadas, pero acopladas de forma débil, en las seis muestras empresariales latinoamericanas. La distinción tiene consecuencias teóricas porque la literatura sobre transformación digital sostiene cada vez con mayor insistencia que el despliegue tecnológico genera valor solo cuando las capacidades organizativas y las rutinas estratégicas se desarrollan en paralelo (Vial, 2019; Verhoef et al., 2021). Los presentes resultados traducen esa proposición en un contraste observable a escala de empresa: una firma puede puntuar alto en presencia web, fiabilidad de internet y transacciones electrónicas, y mostrar al mismo tiempo escasos indicios de seguimiento estructurado del desempeño, horizontes de metas, calibración de esas metas o conocimiento interno de las mismas. La asociación se mantiene, atenuada, cuando se retiran los establecimientos situados en el piso estructural de la medida informacional, lo que descarta que el resultado sea un artificio de la arquitectura de filtros del cuestionario.

El perfil digitalmente avanzado con restricción informacional representa por ello algo más que un grupo de madurez intermedia. Su centroide combina una puntuación digital elevada con una puntuación excepcionalmente baja en gestión de la información, impulsada por la mayoría de establecimientos que no reportan indicadores con seguimiento ni metas de producción o servicio. Identifica, en consecuencia, una configuración en la que la digitalización transaccional o de cara al cliente coexiste con la ausencia declarada de las rutinas directivas internas necesarias para convertir la información operativa en control y aprendizaje organizados. Estos datos no permiten determinar si tales establecimientos gestionan la información por canales informales que el instrumento no capta. Tal lectura resulta compatible con la evidencia de que la capacidad de gestión de la información opera como base de las capacidades de proceso y de desempeño (Mithas et al., 2011), y con la investigación que documenta el fracaso de la implementación de estrategias digitales cuando los mecanismos organizativos no acompañan el cambio tecnológico (Correani et al., 2020). La reestimación sin ceros estructurales matiza esta lectura en un punto importante: el perfil sobrevive con identidad propia, de modo que la configuración es sustantiva, pero su centroide informacional asciende de 10,09 a 38,67. La distancia entre ambas cifras mide exactamente cuánto de la separación aparente procedía de la codificación y cuánto del fenómeno organizativo, y aconseja describir el perfil como una asimetría marcada antes que como un vacío informacional absoluto.

La literatura organizativa ofrece un vocabulario preciso para esta configuración. Bromley y Powell (2012) distinguen el desacoplamiento entre política y práctica, en el que la organización adopta formalmente un dispositivo que no llega a implementar, del desacoplamiento entre medios y fines, en el que la práctica se implementa de manera efectiva pero permanece desvinculada de los resultados que se le atribuyen. El perfil aquí identificado corresponde a la segunda variante: la infraestructura digital opera, procesa pagos y sostiene la relación con el cliente, pero no alimenta el circuito de seguimiento, fijación de metas y difusión interna que convertiría esa operación en aprendizaje directivo. La formulación converge con la explicación canónica de la paradoja de la productividad de las tecnologías de la información, según la cual el rendimiento de la inversión digital depende de complementos organizativos —rediseño de procesos, capacitación, sistemas de medición— cuya construcción es más lenta y más costosa que la adquisición de la tecnología (Brynjolfsson y Hitt, 2000). El desajuste documentado no sería entonces un estadio intermedio de una trayectoria única de modernización, sino la manifestación, a escala de firma, de una asincronía entre la adopción de medios y la construcción de los fines directivos que los justifican. La descomposición en dos partes aporta un indicio congruente: la

capacitación formal multiplica por más de cinco la probabilidad de que exista alguna rutina formalizada, muy por encima de cualquier otro predictor, lo que sitúa el desarrollo de capacidades —y no la dotación tecnológica— en el origen del acoplamiento.

La solución de tres conglomerados tampoco sustenta el supuesto de que la preparación empresarial se organice en cuatro cuadrantes simétricos de valores altos y bajos. No emergió un conglomerado estable de empresas con capacidad informacional y restricción digital en la partición seleccionada por los índices de validez interna, lo cual constituye una ausencia de evidencia en $k = 3$ antes que una demostración de que esa configuración no exista. La reestimación sobre la submuestra sin ceros estructurales refuerza esta cautela: allí el tercer conglomerado combina una puntuación digital baja (57,01) con una informacional próxima al promedio muestral (55,81), esto es, se aproxima a la configuración presuntamente ausente. Lo que la solución principal registra como ausencia responde, al menos en parte, a que la masa de establecimientos sin rutinas declaradas desplaza el centro de gravedad de la partición. Una interpretación plausible es que, entre los establecimientos con veinte trabajadores o más, las rutinas estructuradas de seguimiento y gestión de metas coexisten cada vez más con un nivel mínimo de infraestructura operativa digital. La trayectoria inversa parece más frecuente: las empresas pueden adquirir sitios web, conectividad y capacidad de pago electrónico sin desarrollar rutinas de gestión de la información de estructuración comparable. Esta asimetría debería investigarse de forma longitudinal antes de interpretarse como una secuencia evolutiva.

Los patrones entre países muestran además por qué los promedios regionales resultan insuficientes. Perú, Costa Rica y Colombia concentraron las mayores participaciones de líderes de preparación dual, mientras El Salvador acumuló la mayor proporción de rezagados en ambas dimensiones. Paraguay y México presentaron las proporciones más altas de empresas digitalmente avanzadas con restricción informacional. Estas configuraciones concuerdan con la observación regional más amplia de que la digitalización empresarial latinoamericana avanza bajo condiciones heterogéneas de política, competencias, financiamiento y estructura productiva (OCDE/CAF/SELA, 2024; CEPAL, 2025). Advierten también contra la tentación de interpretar un aumento nacional de la adopción digital visible como prueba de que las capacidades directivas intensivas en información hayan madurado al mismo ritmo.

El caso de Paraguay. Paraguay merece consideración específica por ser el único país cuya asociación entre dominios es negativa y por concentrar simultáneamente la proporción más alta de empresas digitalmente avanzadas con restricción informacional. Las dos observaciones se explican mutuamente: si la digitalización avanza en un segmento de firmas que no ha formalizado sus rutinas de seguimiento, mientras la formalización directiva subsiste en firmas con infraestructura digital más modesta, el ordenamiento conjunto se invierte. Tres verificaciones permiten descartar la hipótesis del artefacto muestral. Primera, el intervalo bootstrap estratificado del coeficiente excluye el cero con holgura ($[-0,484; -0,184]$), de modo que el signo negativo no es atribuible al error de muestreo pese a los 192 establecimientos disponibles. Segunda, la asociación permanece negativa al excluir los establecimientos del piso estructural ($r = -0,268$ sobre 164 casos), lo que descarta que la inversión provenga de la concentración de casos en el origen. Tercera, la proporción de establecimientos con puntuación informacional nula en Paraguay (14,6 %) es inferior a la de El Salvador (23,2 %) y México (18,5 %), países donde la asociación es positiva; la bimodalidad de la medida no distingue al caso paraguayo. El patrón parece, por tanto, sustantivo. La evaluación regional que sitúa a Paraguay en el estrato intermedio de desarrollo digital, con infraestructura y uso empresarial rezagados respecto de los países líderes pero con adopción comparativamente extendida de servicios financieros digitales, ofrece un contexto plausible para una difusión digital concentrada en los flujos de pago y desacoplada de la modernización directiva (Chavarry Galvez y Revinova, 2025). La distinción definitiva entre un patrón institucional y una peculiaridad de la ronda de 2023 no puede resolverse con una sola aplicación del instrumento, y queda planteada como hipótesis contrastable con las rondas sucesivas de la Encuesta de Empresas.

Los resultados del aprendizaje automático explicable afinan esta interpretación sin convertir la asociación en causalidad. El tamaño de la empresa, la capacitación formal, la certificación de calidad y el contexto nacional resultaron consistentemente relevantes para distinguir los perfiles, mientras la exportación y la propiedad extranjera fueron más frecuentes entre los líderes de preparación dual. Tales patrones son coherentes con la investigación que vincula la calidad de la gestión con la escala de la firma y con sus características organizativas (Bloom et al., 2019; Viganola y Diallo, 2025). La capacitación formal merece atención particular porque diferencia a los líderes del grupo digitalmente avanzado con restricción informacional, lo

que refuerza la evidencia latinoamericana previa sobre el papel mediador de las competencias y la formación digitales en la relación entre madurez digital y desarrollo empresarial (Gonzalez-Tamayo et al., 2023). La normalización de las atribuciones por cardinalidad refuerza precisamente esta lectura: una vez neutralizado el efecto aritmético del número de niveles, la capacitación formal y el tamaño encabezan la contribución media por indicador, y el contexto nacional retrocede a posiciones intermedias en dos de los tres perfiles.

El desempeño moderado en validación cruzada y los resultados más débiles en la exclusión sucesiva de países son, a su vez, analíticamente informativos. Un modelo capaz de clasificar los perfiles con exactitud casi perfecta despertaría la sospecha de que el análisis de conglomerados no hizo más que replicar unas pocas variables organizativas evidentes. Las configuraciones de preparación conservan, en cambio, una heterogeneidad no explicada considerable. El deterioro al predecir un país no observado indica que el contexto institucional y de mercado modifica la manera en que características empresariales comunes se relacionan con la preparación. La explicabilidad debería servir, por tanto, al diagnóstico contextual antes que a la elaboración de un ordenamiento universal de empresas o países. Las pruebas de permutación permiten sostener esta lectura con mayor precisión: el desempeño es modesto pero estadísticamente inequívoco tanto dentro de la muestra agrupada como al proyectar sobre países no observados, de manera que la moderación de las cifras refleja la complejidad del fenómeno y no la ausencia de señal.

Este estudio se aparta de los marcos recientes de preparación organizativa para la inteligencia artificial tanto en el constructo como en el propósito. Kalleparambil y Akoum (2025) y Naheed et al. (2025) se concentran en determinar si las empresas reúnen condiciones multidimensionales para adoptar inteligencia artificial, mientras Sriwichien et al. (2026) combinan conglomerados y aprendizaje automático explicable para perfilar esa preparación en la industria logística tailandesa. El análisis aquí presentado evalúa, en cambio, un desajuste multipaís entre la capacidad operativa digital y la práctica operativa de gestión de la información en empresas latinoamericanas. La inteligencia artificial explicable se emplea solo después de derivar los perfiles, con el fin de identificar las características organizativas asociadas a ellos. Esta separación entre la formación del constructo y la explicación del modelo reduce el riesgo de definir la preparación mediante las mismas variables que después se usan para explicarla.

Implicaciones de política diferenciadas por país. La heterogeneidad documentada desaconseja un instrumento único para la región. En El Salvador, donde la brecha negativa proviene de un nivel digital que resulta el más bajo de los seis países en los cuatro sectores examinados, y donde casi la mitad de los establecimientos se clasifican como rezagados en ambas dimensiones, la secuencia razonable comienza por la infraestructura y el costo de la conectividad; la construcción, con el 58,3 % de sus establecimientos sin rutina alguna declarada, exige además un tratamiento sectorial específico. En Perú, con la puntuación informacional ponderada más alta y la mayor participación de líderes duales, el margen de mejora se desplaza hacia la profundidad de los usos digitales y hacia la extensión de las prácticas directivas al segmento que aún no las adopta. En Paraguay y en México, donde predomina el perfil digitalmente avanzado con restricción informacional, el instrumento pertinente no es la subvención a la adopción tecnológica —ya materializada en conectividad y medios de pago— sino la asistencia técnica en sistemas de medición del desempeño, definición de metas y comunicación interna, esto es, los complementos organizativos cuya ausencia el índice detecta. La literatura sobre fomento productivo en la región converge en esta lectura al advertir que los programas de digitalización rinden poco cuando no se articulan con instrumentos de desarrollo de capacidades directivas (Dini et al., 2021; OCDE/CAF/SELA, 2024).

Varias limitaciones delimitan la interpretación. La medida de Preparación Operativa en Gestión de la Información es un indicador indirecto y empírico, basado en prácticas de seguimiento y gobernanza de metas, que no constituye una escala integral validada de gestión empresarial de la información. El Índice de Preparación Digital, por su parte, capta capacidades operativas visibles y no la totalidad de las dimensiones estratégica, cultural, arquitectónica o de ciberseguridad de la transformación digital. El análisis factorial confirmatorio añade una limitación de medida que conviene declarar sin rodeos: mientras la puntuación informacional se comporta como una escala unidimensional con fiabilidad aceptable, los cuatro componentes digitales no comparten un factor común, de manera que el Índice de Preparación Digital solo admite interpretación como compuesto formativo. Su valor resume la extensión de la capacidad operativa digital a través de facetas heterogéneas, pero dos establecimientos con idéntica puntuación pueden diferir por completo en su composición, y la escasa variabilidad de la conectividad en esta población hace que el índice se apoye

de hecho en los flujos de pago y en la presencia web. El diseño es transversal y no permite establecer si la capacitación, la certificación o la internacionalización provocan el tránsito entre perfiles. Los conglomerados son estructuras descriptivas dependientes de los datos, cuyos índices de validez miden reproducibilidad y no la existencia de subpoblaciones discretas, extremo que las especificaciones basadas en densidad confirman de manera directa al resolver la nube en un único conglomerado conexo para radios de vecindad moderados; los valores de Shapley, a su vez, explican la atribución del modelo y no mecanismos causales, y sus magnitudes agrupadas favorecen a los predictores categóricos con más niveles. El piso estructural de la medida de gestión de la información implica que el perfil digitalmente avanzado con restricción informacional queda definido en buena medida por la ausencia declarada de rutinas de seguimiento y metas. El Índice de Preparación Digital presenta además una marcada asimetría negativa, con una mediana de 83,75, lo que limita su poder discriminante entre los establecimientos más digitalizados. La restricción a establecimientos de veinte trabajadores o más se justifica metodológicamente por el módulo de prácticas de gestión del Banco Mundial, pero acota la generalización a las micro y pequeñas empresas. Seis países, finalmente, no pueden representar toda la diversidad institucional de América Latina y el Caribe. Las verificaciones incorporadas en esta versión —exclusión de los ceros estructurales, descomposición en dos partes, análisis factorial confirmatorio, pruebas de permutación, especificaciones alternativas de conglomerados y diagnóstico de colinealidad— delimitan estas limitaciones y cuantifican varias de ellas, pero no las suprimen.

CONCLUSIONES

El análisis conjunto de 2143 establecimientos de seis países latinoamericanos indica que la preparación digital y la preparación operativa en gestión de la información constituyen capacidades organizativas distintas. Su débil asociación positiva muestra que empresas con capacidad operativa digital comparable pueden diferir notablemente en la medida en que dan seguimiento a la información de desempeño, gobiernan sus metas y las difunden internamente. La estructura descriptiva más reproducible estuvo formada por tres perfiles: líderes de preparación dual, empresas digitalmente avanzadas con restricción informacional y empresas rezagadas en ambas dimensiones. Los tres perfiles persisten al excluir los establecimientos situados en el piso estructural de la medida informacional, si bien con una separación menos pronunciada.

Los hallazgos respaldan una implicación directiva: los programas de digitalización no deberían evaluarse únicamente por la adopción de tecnología o por la conectividad transaccional. El desarrollo de la fuerza laboral, la formalización de las prácticas directivas, las rutinas orientadas a la calidad y un diseño organizativo sensible al contexto exigen atención paralela si se espera que las empresas conviertan los recursos digitales en capacidad de decisión basada en información. En el plano operativo, la recomendación con mayor respaldo en estos datos consiste en vincular los instrumentos de apoyo a la adopción digital —subvenciones a la conectividad, a la presencia web o a la incorporación de medios de pago electrónico— a un componente obligatorio de formación directiva en medición del desempeño, definición de metas y comunicación interna de objetivos. La capacitación formal es el predictor organizativo que con mayor nitidez separa a los líderes duales del perfil digitalmente avanzado con restricción informacional, y el que más eleva la probabilidad de que exista alguna rutina formalizada de seguimiento. La secuencia importa: un programa que financie tecnología sin el complemento formativo reproduce, con recursos públicos, el desajuste que este trabajo documenta. Para la investigación, la Brecha de Preparación Digital-Informacional ofrece una herramienta descriptiva transparente para examinar desajustes entre dominios, siempre que no se interprete como un índice universal de madurez. Conviene declararlo sin ambigüedad: la brecha no constituye un índice de ordenamiento entre países. Su signo y su magnitud dependen de la estandarización sobre esta muestra analítica, de suerte que un valor negativo señala una posición digital relativa inferior a la informacional dentro de este conjunto de seis países, y no un déficit absoluto de preparación. Cualquier empleo del indicador para clasificar economías nacionales, o para derivar de esa clasificación asignaciones presupuestarias, carecería de fundamento métrico.

El trabajo futuro debería validar ambas dimensiones con instrumentos específicos de gestión de la información y de transformación digital, extender el análisis a otros países latinoamericanos y a empresas de menor tamaño, y aprovechar las encuestas empresariales repetidas donde la comparabilidad longitudinal permita examinar las transiciones entre perfiles. Un diseño longitudinal resultaría especialmente valioso para determinar si las empresas digitalmente avanzadas con restricción informacional desarrollan después una capacidad estructurada de gestión de la información o permanecen desalineadas de forma persistente. La

construcción de una medida digital cuya dimensionalidad haya sido establecida a priori, y no verificada a posteriori, constituye asimismo una tarea pendiente para la agenda de medición regional.

Disponibilidad de datos y código

Los microdatos originales de las Encuestas de Empresas del Banco Mundial continúan sujetos al protocolo de acceso de dicha institución y están disponibles en el portal de Enterprise Surveys. El archivo de análisis derivado, las salidas estadísticas y las figuras se encuentran disponibles a solicitud del autor de correspondencia y se presentan como material complementario de investigación. El material complementario de esta versión revisada incorpora las rutinas de los análisis añadidos durante la revisión: exclusión de ceros estructurales, especificación en dos partes, análisis factorial confirmatorio sobre matrices policóricas, pruebas de permutación, conglomerados alternativos y diagnóstico de colinealidad.

REFERENCIAS BIBLIOGRÁFICAS

- Banco Mundial. (2026). Enterprise Surveys indicator descriptions. <https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/methodology/Indicator-Description.pdf>
- Bloom, N.; & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4), 1351–1408. <https://doi.org/10.1162/qjec.2007.122.4.1351>
- Bloom, N.; Brynjolfsson, E.; Foster, L.; Jarmin, R.; Patnaik, M.; Saporta-Eksten, I.; & Van Reenen, J. (2019). What drives differences in management practices? *American Economic Review*, 109(5), pp. 1648–1683. <https://doi.org/10.1257/aer.20170491>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bromley, P.; & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *The Academy of Management Annals*, 6(1), 483–530. <https://doi.org/10.1080/19416520.2012.684462>
- Brynjolfsson, E.; & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4), 23–48. <https://doi.org/10.1257/jep.14.4.23>
- Comisión Económica para América Latina y el Caribe. (2025). *Overcoming development traps in Latin America and the Caribbean in the digital age: The transformative potential of digital technologies and artificial intelligence* (LC/CMSI.9/3/Rev.1). CEPAL. <https://hdl.handle.net/11362/81383>
- Chavarry Galvez, D. P.; & Revinova, S. (2025). Assessing digital technology development in Latin American countries: Challenges, drivers, and future directions. *Digital*, 5(2), 20. <https://doi.org/10.3390/digital5020020>
- Chen, T.; & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. En *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Correani, A.; De Massis, A.; Frattini, F.; Messeni Petruzzelli, A.; & Natalicchio, A. (2020). Implementing a digital strategy: Learning from the experience of three digital transformation projects. *California Management Review*, 62(4), pp. 37–56. <https://doi.org/10.1177/0008125620934864>
- Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39(5), pp. 829–844. <https://doi.org/10.2307/1909582>
- Diamantopoulos, A.; & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, 38(2), 269–277. <https://doi.org/10.1509/jmkr.38.2.269.18845>
- Dini, M.; Gligo, N.; & Patiño, A. (2021). *Transformación digital de las mipymes: elementos para el diseño de políticas* (LC/TS.2021/99). CEPAL. <https://hdl.handle.net/11362/47183>

- Ester, M.; Kriegel, H.-P.; Sander, J.; & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. En Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (pp. 226–231). AAAI Press.
- Fox, J.; & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87(417), 178–183. <https://doi.org/10.1080/01621459.1992.10475190>
- Fraley, C.; & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458), pp. 611–631. <https://doi.org/10.1198/016214502760047131>
- Gonzalez-Tamayo, L. A.; Maheshwari, G.; Bonomo-Odizzio, A.; Herrera-Avilés, M.; & Krauss-Delorme, C. (2023). *Factors influencing small and medium size enterprises development and digital maturity in Latin America*. *Journal of Open Innovation: Technology, Market, and Complexity*, 9(2), 100069. <https://doi.org/10.1016/j.joitmc.2023.100069>
- Grupo Banco Mundial. (2025). World Bank Enterprise Survey 2023: Colombia, Costa Rica, El Salvador, Mexico, Paraguay, and Peru [conjuntos de datos]. World Bank Microdata Library. https://microdata.worldbank.org/collections/enterprise_surveys
- Hausmann, V.; Williams, S. P.; Hardy, C. A.; & Schubert, P. (2014). Enterprise information management readiness: A survey of current issues, challenges and strategy. *Procedia Technology*, 16, 42–51. <https://doi.org/10.1016/j.protcy.2014.10.066>
- Hubert, L.; & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193–218. <https://doi.org/10.1007/BF01908075>
- Kalleparambil, S. A.; & Akoum, M. (2025). AI organisational readiness for SMEs: A tailored model for AI adoption success. *Journal of Information & Knowledge Management*, 24(6), 1–21. <https://doi.org/10.1142/S0219649225500765>
- Kettinger, W. J.; Ryoo, S. Y.; & Marchand, D. A. (2021). We're engaged! Following the path to a successful information management capability. *The Journal of Strategic Information Systems*, 30(3), 101681. <https://doi.org/10.1016/j.jsis.2021.101681>
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Lundberg, S. M.; & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. En Advances in Neural Information Processing Systems 30 (pp. 4765–4774).
- Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, 2, pp. 56–67. <https://doi.org/10.1038/s42256-019-0138-9>
- Mithas, S.; Ramasubbu, N.; & Sambamurthy, V. (2011). How information management capability influences firm performance. *MIS Quarterly*, 35(1), 237–256. <https://doi.org/10.2307/23043496>
- Naheed, S.; Pinto, R.; & Pirola, F. (2025). *A preliminary multidimensional AI readiness assessment model for SME's*. *Procedia Computer Science*, 253, pp. 774–783. <https://doi.org/10.1016/j.procs.2025.01.139>
- OCDE/CAF/SELA. (2024). SME policy index: Latin America and the Caribbean 2024: Towards an inclusive, resilient, and sustainable recovery. OECD Publishing. <https://doi.org/10.1787/ba028c1d-en>
- Ojala, M.; & Garriga, G. C. (2010). Permutation tests for studying classifier performance. *Journal of Machine Learning Research*, 11, 1833–1863.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pp. 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

- Shafiabady, N.; Hadjinicolaou, N.; Hettikankanamage, N.; MohammadiSavadkoohi, E.; Wu, R. M. X.; & Vakilian, J. (2024). eXplainable Artificial Intelligence (XAI) for improving organisational regility. PLOS ONE, 19(4), e0301429. <https://doi.org/10.1371/journal.pone.0301429>
- Sriwichien, W.; Narkbunnum, W.; & Wisaeng, K. (2026). Profiling organizational AI readiness in Thailand's logistics industry using TOE-UTAUT features, clustering analysis, and explainable machine learning. Information, 17(7), 672. <https://doi.org/10.3390/info17070672>
- Verhoef, P. C.; Broekhuizen, T.; Bart, Y.; Bhattacharya, A.; Dong, J. Q.; Fabian, N.; & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889–901. <https://doi.org/10.1016/j.jbusres.2019.09.022>
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), + 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- Viganola, D.; & Diallo, M. Y. (2025). How management practices differ in the EU-27 (Enterprise Note No. 42). Banco Mundial. <https://doi.org/10.1596/42897>