RAES
Revista Argentina
de Educación Superior

Transiciones tecnológicas y límites de la inferencia en educación superior: deriva semántica de métricas, sesgos de atribución y arquitectura de medición resiliente.

Technological Transitions and the Limits of Inference in Higher Education: Semantic Drift of Metrics, Attribution Biases, and a Resilient Measurement Architecture.

Por Hugo Roger PAZ[1]

Roger Paz, H. (2026). Transiciones tecnológicas y límites de la inferencia en educación superior: deriva semántica de métricas, sesgos de atribución y arquitectura de medición resiliente. *Revista RAES*, XVIII(33), pp. 13-31.

Resumen

En educación superior, los indicadores cuantitativos de desempeño académico sustentan decisiones de política institucional, asignación de recursos y evaluación de calidad. Su validez inferencial, sin embargo, se vuelve frágil durante transiciones tecnológicas que reconfiguran reglas, incentivos, restricciones y prácticas (por ejemplo, virtualización acelerada, digitalización de procesos, cambios en modalidades de evaluación y en sistemas de registro).

Este artículo analiza un problema recurrente: inferencias estadísticamente "correctas" pueden resultar sistemáticamente engañosas cuando cambia el proceso que genera los datos, aun si la fórmula del indicador permanece constante. Integrando teoría de sistemas adaptativos complejos, sociología de la cuantificación y teoría de la medición, se describen patrones típicos de inestabilidad inferencial bajo transición: cambios de nivel, quiebres de tendencia, heterogeneidad creciente entre cohortes y segmentos, sustitución de esfuerzo y deriva semántica de métricas aparentemente estables.

Como contribución aplicada, se propone una arquitectura mínima de medición resiliente basada en contrato de medición, chequeos de invariancia, panel complementario de indicadores, detección de cambios de régimen, triangulación de fuentes y gobernanza de métricas. El objetivo es reducir errores de diagnóstico, distinguir cambios reales de aprendizaje de artefactos inducidos por transición y orientar evaluaciones más robustas, transparentes y socialmente responsables.

Palabras Clave educación superior / indicadores de desempeño / medición educativa / deriva semántica / analítica del aprendizaje.

[1] Universidad Nacional de Tucumán, Argentina/ hpaz@herrera.unt.edu.ar / http://orcid.org/0000-0002-6291-2545

Abstract

In higher education, quantitative performance indicators underpin institutional policy decisions, resource allocation, and quality evaluation. Their inferential validity, however, becomes fragile during technological transitions that reconfigure rules, incentives, constraints, and practices (e.g., rapid virtualisation, digitalisation of processes, shifts in assessment modalities, and changes in record-keeping systems).

This paper analyses a recurrent problem: statistically "correct" inferences may become systematically misleading when the data-generating process changes, even if the indicator's formula remains unchanged. By integrating complex adaptive systems theory, the sociology of quantification, and measurement theory, the paper describes typical patterns of inferential instability under transition: level shifts, trend breaks, increasing heterogeneity across cohorts and segments, effort substitution, and semantic drift in seemingly stable metrics. As an applied contribution, the paper proposes a minimal resilient measurement architecture based on a measurement contract, invariance checks, a complementary indicator panel, regime-shift detection, triangulation of sources, and metric governance. The goal is to reduce diagnostic errors, distinguish genuine learning changes from transition-induced artefacts, and support more robust, transparent, and socially responsible evaluation.

Key words higher education / performance indicators / educational measurement / semantic drift / learning analytics.

Planteamiento del problema o tema objeto de estudio

Medir para decidir: la racionalidad de las métricas y su fragilidad

Lejos de ser un ejercicio metodológico aislado, la medición en la educación superior actual funciona como una infraestructura esencial para el gobierno del sistema. El uso de indicadores de rendimiento, regularidad, no presentación y analíticas de aprendizaje permite a las instituciones asignar presupuestos, rediseñar planes de estudio y evaluar el impacto de sus políticas de calidad. Esa racionalidad es comprensible: los sistemas universitarios son grandes, costosos, políticamente sensibles y temporalmente largos; sin medición, se depende de juicios ad hoc. También obliga a preguntarse qué dimensiones de la educación se están privilegiando cuando una institución gobierna a través de números (Biesta, 2010).

El problema es que esa medición suele tratarse como si fuera un termómetro: si sube o baja, "algo real" sube o baja. En sistemas sociales, sin embargo, los indicadores no son termómetros: son parte del sistema. Una vez que una métrica entra en el circuito de decisiones, cambia incentivos, cambia conductas, cambia estrategias, y por lo tanto cambia su propio significado. Esta es la intuición compartida por la Ley de Goodhart y el principio de Campbell: cuando un indicador se convierte en objetivo, tiende a degradarse como medida del constructo que pretendía capturar. En educación superior, ese riesgo se amplifica porque la vigilancia métrica es densa, los incentivos son múltiples y los actores (estudiantes, cátedras, gestión) son adaptativos.

Qué se entiende por transición tecnológica (y por qué importa)

Este trabajo se centra en un tipo de perturbación que altera el "régimen" del sistema: la transición tecnológica asociada a plataformas digitales, conectividad ubicua y mediación algorítmica (LMS, repositorios, buscadores, IA generativa, evaluación online, comunicación asíncrona). No se asume determinismo tecnológico ni un vínculo causal simple del tipo "apareció X → cayó Y". La tecnología no opera como una fuerza única que empuja resultados; opera reconfigurando la **estructura de oportunidades**: reduce costes marginales de acceso a información, habilita nuevas secuencias y ritmos de estudio, introduce mecanismos alternativos de señalización del esfuerzo y del aprendizaje, y permite nuevas formas de coordinación (o evasión) de reglas.

Desde una perspectiva de sistemas adaptativos complejos, estos cambios se parecen más a un choque estructural exógeno que modifica condiciones de contorno (costes, tiempos, opciones disponibles) que a una reforma interna del currículo. Los agentes siguen orientados a objetivos, siguen adaptándose, pero lo hacen en un ambiente distinto. Y cuando cambia el ambiente, cambia el mapa de estrategias plausibles. En consecuencia, un mismo indicador puede estar midiendo "otra cosa" sin que el modelo estadístico haya cometido ningún error. Leído desde enfoques complejistas del currículo, esto implica que la relación entre reglas, enseñanza, evaluación y aprendizaje deja de ser lineal y pasa a depender de interacciones emergentes y no totalmente controlables (Gough, 2013; Osberg & Biesta, 2010).

La paradoja central: inferencias "correctas" que se vuelven engañosas

La investigación educativa ha elevado notablemente sus estándares analíticos: diseños longitudinales, inferencia causal (cuando se puede), control de confusores, modelos jerárquicos, verificación de robustez, cuantificación de incertidumbre. Muchas inferencias, en términos técnicos, son correctas.

La paradoja es que **la exactitud estadística no garantiza la validez sustantiva**. En múltiples contextos universitarios aparecen patrones empíricamente robustos (descensos de aprobación, aumentos de no participación en evaluaciones, cambios de pendiente en la progresión, reconfiguración de la varianza entre cohortes) que se interpretan de manera casi automática como deterioro de capacidades estudiantiles, caída de estándares o fracaso institucional. Esas interpretaciones suelen persistir incluso cuando no se observa una degradación proporcional en la estructura curricular, en los recursos docentes o en el perfil de ingreso.

Este artículo sostiene que una parte relevante de esas lecturas no se explica principalmente por “malos datos” o “modelos débiles”, sino por un fenómeno más incómodo: **inestabilidad inferencial**. En sistemas sociales adaptativos, la relación entre indicadores observables y constructos subyacentes (aprendizaje, compromiso, esfuerzo, dificultad efectiva, calidad) puede desacoplarse cuando el sistema atraviesa un cambio de régimen. Bajo transición tecnológica, las métricas pueden conservar su coherencia interna (seguir “funcionando” estadísticamente) mientras pierden su significado externo: lo que antes implicaba “bajo aprendizaje” puede ahora implicar “nueva estrategia”; lo que antes era “falta de esfuerzo” puede convertirse en “sustitución eficiente de esfuerzo”; lo que antes era “riesgo” puede ser “optimización racional del tiempo”.

En términos de teoría social, esto equivale a decir que la medición está atravesada por **reflexividad**: los actores aprenden, responden y ajustan su conducta al entorno métrico y tecnológico. En términos de economía política de la educación, es una versión operacional de la Crítica de Lucas: los parámetros conductuales que vinculan una política con un resultado cambian cuando la política (o el entorno) cambia. En términos de complejidad, el sistema cambia su atractor: se estabiliza en otra configuración de comportamientos agregados, y las series temporales capturan ese desplazamiento. El error, entonces, no es de cálculo; es de **semántica**.

Relevancia práctica: por qué no es un asunto “técnico”

Este problema no es un detalle filosófico para el final del manuscrito. Es un riesgo práctico de alto impacto porque las métricas gobiernan decisiones reales: cupos, asignación de tutorías, reformas de cátedra, cambios reglamentarios, acreditación, financiamiento, legitimidad institucional. Una interpretación sustantiva errónea puede producir dos daños simultáneos:

1. Daño distributivo: intervenciones “bien intencionadas” que terminan castigando a grupos específicos (por ejemplo, cohortes que adoptan estrategias nuevas de estudio o de rendición bajo ambientes digitales), o que trasladan costos a quienes ya operan con menor margen de maniobra.

2. Daño epistémico y de confianza: cuando la institución interpreta como “declive” lo que en realidad es cambio de estrategia bajo nuevas condiciones, se erosiona la credibilidad de la evaluación, se genera conflicto normativo, y se alimentan respuestas adaptativas defensivas (más optimización, más simulación, menos cooperación).

Por eso, tratar el asunto como un ajuste “técnico” (mejorar un modelo, agregar covariables, “tunear” umbrales) suele ser insuficiente. El núcleo del problema es estructural: cómo se sostiene la validez de la medición cuando el sistema cambia de régimen. En otras palabras: la pregunta no es solo “qué dicen los datos”, sino “qué están diciendo ahora los mismos datos en un mundo distinto”.

Objetivo y contribución del artículo

El objetivo general es construir un marco interpretativo, desde sociología de la medición y teoría de sistemas adaptativos complejos, para explicar por qué pueden observarse patrones longitudinales estadísticamente sólidos que, bajo transición tecnológica, inducen interpretaciones sustantivas equivocadas.

De manera específica, este artículo:

- (i) caracteriza el fenómeno de **deriva semántica** de indicadores educativos bajo cambio de régimen, distinguiendo validez interna de significado externo;

- (ii) identifica firmas típicas de inestabilidad inferencial observables en series longitudinales (cambios de nivel, quiebres de tendencia, reconfiguración de varianza y heterogeneidad entre cohortes) y discute su interpretación plausible bajo adaptación estratégica;

- (iii) propone un conjunto mínimo de criterios para una medición "resiliente a régimen" en contextos universitarios, orientada a reducir el salto ilegítimo desde regularidades empíricas hacia narrativas de declive o déficit.

La contribución principal es conceptual y aplicada: no se discute la necesidad de medir, sino la necesidad de medir con conciencia de régimen. El argumento final es simple y antipático (por eso sirve): en universidades sometidas a transiciones tecnológicas, la inferencia correcta puede ser necesaria y aun así no ser suficiente. La solución no es medir menos, sino medir mejor: con humildad inferencial, trazabilidad de supuestos y mecanismos explícitos para detectar cuándo una métrica dejó de significar lo que creíamos que significaba. A diferencia de trabajos que discuten por separado auditoría, analítica del aprendizaje o IA generativa, este artículo formaliza el mecanismo que enlaza cambio de régimen, reactividad y deriva semántica, y lo traduce en un diseño mínimo de medición institucional.

Antecedentes y fundamentación teórica

Gobernanza por números: cuantificación, auditoría y efectos organizacionales

La expansión de métricas en educación superior no es un fenómeno puramente técnico, sino un cambio en el modo de gobierno institucional. En la medida en que los indicadores se incorporan a la gestión, la acreditación y la asignación de recursos, pasan de "describir" a organizar decisiones, estabilizar categorías y producir comparabilidad entre programas, cohortes y unidades académicas. En este sentido, la cuantificación funciona como una infraestructura socio-técnica: define qué se considera desempeño, riesgo o calidad y, por lo tanto, orienta conductas y prioridades (Espeland & Stevens, 2008). En este sentido, la expansión de tableros, rankings y rutinas de rendición de cuentas puede leerse también como una profundización de lo que Muller (2018) denominó la tiranía de las métricas.

Esta lógica se refuerza en contextos de auditoría y rendición de cuentas. Los procesos de verificación tienden a favorecer aquello que puede registrarse y reportarse con facilidad, promoviendo dinámicas de cumplimiento formal y desplazamiento de fines: se optimiza lo medible aunque ello no represente necesariamente lo más relevante (Power, 1997). En educación superior, esto puede traducirse en un gobierno por tableros donde se confunde estabilidad numérica con estabilidad sustantiva.

Desde una perspectiva de sistemas adaptativos complejos, el punto crítico es que los indicadores no son externos al sistema universitario. Al entrar en circuitos de decisión, se integran en bucles de retroalimentación: medición → interpretación → decisión → adaptación → nueva medición. En ese marco, es esperable que la conducta agregada cambie no solo por transformaciones "reales" del aprendizaje, sino por cambios en incentivos, reglas y posibilidades técnicas. Por ello, las métricas deben tratarse como artefactos sociales y organizacionales: útiles, pero susceptibles de producir efectos secundarios y de cambiar su significado cuando cambia el régimen de funcionamiento. Esta problemática se ha vuelto más visible con la rápida institucionalización de lineamientos sobre IA generativa, recursos universitarios de uso de ChatGPT y formas emergentes de gobernanza algorítmica, que hacen explícita la dimensión político-organizacional de la infraestructura de datos (Wang et al., 2024; Yan et al., 2024; McConvey & Guha, 2024).

Reactividad del indicador bajo presión decisional (principio Goodhart–Campbell)

Un antecedente clave para entender la fragilidad interpretativa de las métricas es el principio de **reactividad del indicador**: cuando una medida se utiliza intensamente para gobernar decisiones, se vuelve endógena, porque los actores ajustan su conducta a la forma en que serán evaluados. Este principio integra dos advertencias clásicas: cuando un indicador se convierte en objetivo, pierde capacidad para medir el constructo que pretendía representar (Goodhart, 1975); y cuanto más se usa un indicador cuantitativo para decisiones sociales, más tiende a distorsionarse y a distorsionar el proceso que monitorea (Campbell, 1979).

En educación superior, esta reactividad no requiere fraude ni intencionalidad explícita. Puede emerger como adaptación racional frente a incentivos institucionales y restricciones reales: cambios en estrategias de rendición, reordenamiento del tiempo de estudio, selección de asignaturas, o modificaciones docentes en instrumentos y criterios de evaluación. Bajo transición tecnológica, la reactividad tiende a intensificarse porque disminuyen costos de acceso a información y herramientas, aumenta la velocidad de difusión de estrategias, y se amplía el repertorio de sustitución de esfuerzo (por ejemplo, repositorios digitales, automatización parcial e IA generativa). En consecuencia, un mismo valor de un indicador puede corresponder a procesos distintos en regímenes distintos. Esto vuelve riesgoso la interpretación automática de variaciones longitudinales como evidencia directa de "declive" o "mejora" del aprendizaje.

Cambio de régimen e inestabilidad semántica: límites de la comparación longitudinal (Lucas–Meredith)

El núcleo del problema analítico puede expresarse en una conclusión operativa: bajo cambio de régimen, los parámetros conductuales no son constantes y la estabilidad técnica no garantiza estabilidad semántica. Cuando cambian reglas, tecnología o condiciones de contorno, los actores ajustan expectativas y estrategias; por lo tanto, las relaciones que conectan políticas, métricas y resultados pueden reconfigurarse. Esta idea se alinea con la advertencia de que los modelos que describen un régimen pueden perder validez al trasladarse a otro, porque el comportamiento del sistema cambia con el entorno (Lucas, 1976).

En paralelo, la investigación longitudinal presupone algún tipo de invariancia: que el indicador mantiene un vínculo estable con el constructo a través del tiempo. Cuando esa invariancia falla, comparar periodos puede equivaler a comparar entidades distintas. Desde la tradición de invariancia en medición, la conclusión relevante para este estudio es que diferencias observadas entre cohortes o periodos no pueden atribuirse automáticamente a cambios del constructo si cambió el régimen de medición o las condiciones que generan el indicador (Meredith, 1993). En términos prácticos, esto significa que un indicador puede mostrar coherencia estadística —ser estable, modelable, predictivo dentro de un periodo— y aun así haber cambiado de significado por la transformación del contexto socio-técnico.

Esta combinación (reactividad + cambio de régimen) permite conceptualizar la **deriva semántica**: el desplazamiento del significado operativo de un indicador sin modificación formal de su definición. En universidades atravesadas por transiciones tecnológicas, este riesgo es especialmente alto, porque se reconfiguran costos, oportunidades y formas de participación académica. Por ello, el análisis de métricas longitudinales debería incorporar procedimientos explícitos para distinguir cambios en el fenómeno de cambios en el significado de la medición.

Síntesis del marco: por qué una métrica "correcta" puede inducir a error

Integrando los antecedentes, el marco teórico del manuscrito sostiene cuatro premisas:

1. La cuantificación en educación superior funciona como **gobernanza por números**: produce comparabilidad y orienta decisiones, con efectos organizacionales propios (Espeland & Stevens, 2008; Power, 1997).
2. Bajo presión decisional, los indicadores exhiben **reactividad**: los actores se adaptan y el indicador puede degradarse como indicador sustitutivo del constructo (Campbell, 1979; Goodhart, 1975).
3. Las transiciones tecnológicas pueden operar como **cambios de régimen**: modifican el espacio de estrategias disponibles y, por lo tanto, alteran los vínculos entre métricas y resultados (Lucas, 1976).
4. La comparación longitudinal presupone invariancia; cuando esa condición se debilita, la estabilidad técnica no basta: puede existir **inestabilidad semántica** (Meredith, 1993).

Sobre esta base, la tesis del manuscrito se vuelve precisa: en contextos universitarios bajo transición tecnológica, es plausible observar series longitudinales estadísticamente consistentes que, sin embargo, inducen interpretaciones sustantivas erróneas si se aplican categorías del régimen anterior. Esta fundamentación abre paso

a los apartados siguientes, donde se describen señales empíricas compatibles con ruptura inferencial y se propone una arquitectura mínima de medición resiliente para la toma de decisiones institucionales.

Metodología

Enfoque y diseño

Este estudio adopta un **diseño cualitativo-conceptual** con orientación socio-técnica y de sistemas adaptativos. El objetivo no es estimar un efecto causal puntual, sino **formalizar el mecanismo** por el cual, bajo transiciones tecnológicas, indicadores longitudinales pueden conservar consistencia estadística y aun así volverse **semánticamente inestables**. El análisis se estructuró como una síntesis guiada por patrones, combinando **abducción** (proponer la explicación más plausible para regularidades observadas) y **codificación de firmas** (identificación de señales recurrentes compatibles con cambio de régimen) (Timmermans & Tavory, 2012).

Procedimiento analítico (abducción + codificación de firmas)

El trabajo se desarrolló en dos movimientos complementarios:

1. Abducción mecanicista. A partir de antecedentes sobre gobernanza por números, reactividad del indicador y cambio de régimen, se construyó una explicación integrada de por qué puede producirse una brecha entre validez técnica y validez sustantiva de las métricas. La abducción se utilizó para articular el marco conceptual con observaciones empíricas reportadas en la literatura, buscando consistencia con agencia adaptativa y retroalimentación institucional.
2. Codificación de firmas. Sobre la base de reportes longitudinales en educación superior bajo digitalización/virtualización/IA, se identificaron firmas observables compatibles con ruptura inferencial, utilizadas como alertas interpretativas (no como pruebas causales). Estas firmas se consolidaron en una taxonomía de patrones para orientar la discusión y la propuesta aplicada.

La secuencia metodológica (abducción y codificación de firmas) y la correspondencia entre lógica, insumos y productos del análisis se resumen en la Tabla 1.

Tabla 1. Núcleo metodológico: lógica, insumos y productos del análisis

Etapa	Lógica	Insumos principales	Producto
A. Abducción	Explicación más plausible para regularidades bajo transición tecnológica	Literatura de cuantificación y auditoría; reactividad del indicador; cambio de régimen; invariancia/estabilidad semántica	Modelo conceptual de "ruptura inferencial" y deriva semántica
B. Codificación de firmas	Identificación de señales recurrentes en series longitudinales	Estudios/revisiones con resultados longitudinales y discusiones de validez/transferencia	Taxonomía de firmas: (i) cambios de nivel, (ii) quiebres de tendencia, (iii) reconfiguración de varianza/heterogeneidad, (iv) no presentación/ausentismo estratégico, (v) desacople entre indicadores
C. Integración aplicada	Traducir marco + firmas a un mínimo institucional operable	Resultados A+B	Arquitectura mínima de medición resiliente (contrato, bitácora, checks semánticos, trazabilidad y gobernanza interpretativa)

Fuente: elaboración propia.

Criterio de selección de fuentes

También se incorporó literatura reciente sobre IA generativa y gobernanza algorítmica en educación superior para conectar la discusión de cambio de régimen con la capa normativa contemporánea.

Se priorizaron trabajos que: (i) conceptualizan el rol de métricas como instrumentos de gobierno; (ii) discuten reactividad, incentivos o distorsiones asociadas al uso decisional de indicadores; y/o (iii) analizan comparabilidad longitudinal bajo cambios contextuales (digitalización, virtualización, cambios de reglas). Se excluyeron aportes meramente instrumentales (p. ej., modelos predictivos sin discusión de validez interpretativa) o reportes sin trazabilidad metodológica suficiente.

Salida metodológica: qué valida este diseño

Este diseño valida **coherencia mecanicista** y **utilidad operacional**: (i) que la explicación propuesta sea consistente con agencia adaptativa y retroalimentación institucional, y (ii) que las firmas identificadas puedan funcionar como señales prácticas de alerta semántica para evaluación institucional y políticas universitarias. En este sentido, el aporte del estudio es principalmente **conceptual-aplicado**: ofrece un marco y una arquitectura mínima para reducir errores de interpretación bajo cambio de régimen, más que estimaciones de magnitud o causalidad.

Resultados: señales empíricas de ruptura inferencial (versión reformateada)

Enfoque del apartado: "firmas" como alertas de cambio de régimen

Los resultados se presentan como **firmas observables** recurrentes en series longitudinales de educación superior en contextos de transición tecnológica (digitalización, virtualización, cambio de plataformas y, más recientemente, IA generativa). Estas firmas no se interpretan como pruebas causales de deterioro o mejora, sino como **señales de posible ruptura inferencial**: situaciones en las que el indicador puede seguir siendo estadísticamente coherente, pero haber cambiado su significado operativo por reactividad, adaptación estratégica y cambio de régimen (Campbell, 1979; Espeland & Stevens, 2008; Lucas, 1976; Power, 1997).

Síntesis de patrones (P1–P4): firmas observables, mecanismos de deriva y chequeos mínimos

Las series longitudinales en educación superior bajo transición tecnológica suelen exhibir **cuatro firmas** cuya utilidad principal no es "demostrar" declive o mejora, sino **activar una alerta interpretativa**: revisar si el indicador conserva el mismo significado operativo o si está capturando conductas distintas bajo un régimen distinto. En primer lugar, el **cambio de nivel (P1)** aparece como un salto abrupto en aprobación, regularidad o ausentismo, típicamente cercano a un hito institucional (cambio de modalidad, LMS o reglas). Este patrón es compatible con una combinación de **cambio de régimen + reactividad**, donde el mismo indicador comienza a agregarse desde conductas distintas (recalibración del riesgo, modificación de instrumentos o cambio en la composición de quienes rinden). El chequeo mínimo recomendado es verificar la **alineación temporal** con la bitácora de régimen y evaluar si se alteran **relaciones históricas** entre indicadores (por ejemplo, asistencia↔aprobación o parciales↔finales).

En segundo lugar, el **quiebre de tendencia (P2)** se expresa como un cambio de pendiente: procesos que mejoraban se estancan o se revierten (o viceversa). En sistemas adaptativos, esto es compatible con el **aprendizaje del sistema**, donde los actores adoptan gradualmente nuevas estrategias, ajustan expectativas y desplazan cuellos de botella. El chequeo mínimo consiste en comparar pendientes entre **materias núcleo y periféricas** y monitorear si simultáneamente crece la **heterogeneidad** (aumento de varianza), lo que suele acompañar la diversificación estratégica.

En tercer lugar, la **reconfiguración de varianza/heterogeneidad (P3)** se observa cuando aumenta la dispersión entre cohortes o subpoblaciones. Este patrón es consistente con **diversificación estratégica** en un nuevo espacio de posibilidades y con redistribución de ventajas socio-técnicas (acceso, alfabetización digital, tiempo disponible),

generando sensibilidad diferencial a reglas y plataformas. El control mínimo no debe centrarse en promedios, sino en las **colas** de la distribución, la segmentación por condiciones materiales (trabajo, conectividad) y la comparación entre estabilidad "intra-grupo" versus divergencia "entre-grupos".

Por último, el **desacople entre indicadores (P4)** —cuando relaciones históricas se rompen, por ejemplo, actividad en LMS que deja de predecir aprobación— constituye la señal más directa de posible **deriva semántica**: el indicador puede conservar estabilidad interna pero perder validez externa por reactividad y reconfiguración de prácticas. El chequeo mínimo recomendado es monitorear una **matriz de relaciones** (correlaciones/elasticidades) comparando periodos pre/post régimen y analizar por periodos separados, evitando extrapolar interpretaciones del régimen anterior al nuevo.

En conjunto, estos patrones deben tratarse como **alertas y no veredictos**: si aparecen, la respuesta institucional correcta no es concluir inmediatamente "éxito" o "declive", sino activar revisión de régimen y comprobar si cambió el significado del indicador antes de sostener inferencias sustantivas o decisiones de alto impacto.

No presentación y ausentismo estratégico: el patrón que más se moraliza y menos se entiende

Entre los patrones anteriores, el **no presentación/ausentismo** (inscribirse y no rendir; ausentarse de exámenes; sostener regularidad sin presentarse; postergar intentos) merece un tratamiento separado porque suele ser el punto donde la institución cae más rápido en interpretaciones moralizantes ("falta de compromiso", "desinterés", "baja cultura del esfuerzo"). Desde un enfoque sociológico y de complejidad, esa lectura es frecuentemente pobre, porque confunde conducta observada con disposición moral, e ignora que la decisión de rendir es un **cálculo situado**.

El no presentación como decisión bajo incertidumbre

Presentarse a una instancia evaluativa no es un acto "neutral": tiene costos (tiempo, traslado, estrés, exposición pública), riesgos (reprobar, perder regularidad, consecuencias administrativas) y retornos (avanzar, liberar carga futura). Cuando el entorno cambia —por tecnología, modalidad, reglas, disponibilidad de materiales, existencia de recuperatorios o cambios de instrumentos— cambia el cálculo racional.

En transiciones tecnológicas, tres transformaciones son típicas:

1. Reducción de fricciones y expansión de recursos disponibles. Con plataformas, repositorios y herramientas digitales, la sensación de "puedo preparar mejor después" se vuelve más plausible. Esto puede aumentar postergaciones sin que ello implique menor capacidad.

2. Recalibración del riesgo evaluativo. Cambios en modalidad (online/presencial), en controles (antifraude), en instrumentos (más integradores, menos previsibles), o en reglas (recuperatorios, promoción) pueden modificar la percepción de riesgo. En ese contexto, ausentarse puede ser una estrategia de **evitación racional de pérdidas**: no "quemo un intento" cuando la probabilidad de fallo es alta.

3. Optimización del portafolio académico. Estudiantes con restricciones fuertes (trabajo, familia, conectividad) tienden a asignar esfuerzo donde el retorno percibido es mayor o el riesgo es menor. El no presentación puede ser la huella cuantitativa de esa optimización: el sistema se vuelve más selectivo en qué intentos se ejecutan.

En síntesis: el no presentación es un indicador altamente expuesto a **reactividad del sistema**. Puede aumentar por deterioro real, sí. Pero también puede aumentar por **cambio de estrategia** bajo un régimen nuevo. Por eso es un indicador peligroso para "explicar" y útil para **alertar**.

Mecanismo de deriva: cuando el ausentismo deja de significar lo mismo

El corazón del asunto es semántico: en un régimen, "no rendir" puede significar desconexión; en otro, puede significar gestión de riesgo o reordenamiento temporal. Bajo presión decisional, además, el no presentación puede

volverse endógeno: si la institución sanciona o estigmatiza, los estudiantes ajustan la conducta (p. ej., se inscriben menos, o rinden menos en condiciones subóptimas), y el indicador cambia nuevamente de significado. Este es un caso aplicado del principio Goodhart–Campbell: la métrica se convierte en objetivo y el sistema responde (Campbell, 1979; Goodhart, 1975).

Síntesis de resultados: qué aportan estas firmas

El aporte de esta sección es convertir una intuición frecuente ("los números cambiaron y no sabemos qué significa") en un conjunto de **señales operativas**. Cuando aparecen cambios de nivel, quiebres de tendencia, heterogeneidad creciente o desacoples, la institución debería activar una regla simple: **antes de atribuir declive o éxito, auditar si cambió el régimen y el significado**. En particular, el no presentación debe tratarse como un indicador de alta reactividad: útil para vigilancia interpretativa, pero peligroso como base única de diagnósticos morales o decisiones punitivas.

Discusión (integrada con implicaciones para evaluación institucional, calidad y política universitaria)

Relectura del hallazgo central: el error típico no es de cálculo, sino de interpretación

Las firmas empíricas presentadas (cambios de nivel, quiebres de tendencia, heterogeneidad creciente, desacoples entre indicadores y, en particular, no presentación/ausentismo estratégico) convergen en un punto: bajo transición tecnológica, puede existir una brecha entre **validez técnica** (coherencia estadística del indicador dentro de un periodo) y **validez sustantiva** (estabilidad de significado del indicador como indicador sustitutivo del constructo). En un sistema universitario entendido como sistema adaptativo, la medición ingresa en bucles de retroalimentación y, por lo tanto, modifica conductas; este rasgo hace que el problema dominante sea la **ruptura inferencial**: interpretar con categorías del régimen anterior datos producidos por un régimen nuevo.

La consecuencia metodológica es directa: cuando aparecen firmas compatibles con cambio de régimen, la pregunta institucional relevante deja de ser "¿subió o bajó el indicador?" y pasa a ser "¿sigue significando lo mismo?". Esto no niega la posibilidad de deterioro real; niega el atajo interpretativo.

Mecanismo integrado: reactividad + cambio de régimen (por qué el indicador puede "mentir" sin mentir)

La explicación más parsimoniosa de la ruptura inferencial combina dos mecanismos complementarios:

1. Reactividad del indicador bajo presión decisional (Goodhart–Campbell). Cuando una métrica se utiliza para gobernar decisiones (acreditación, metas, evaluación interna), se vuelve endógena: los actores ajustan conducta a lo que se mide. Esto puede distorsionar el indicador sin necesidad de fraude; basta con adaptación racional y aprendizaje del sistema.
2. Cambio de régimen (Lucas + no-invariancia). Si cambian reglas, tecnología o condiciones de contorno, cambian los parámetros conductuales y el vínculo indicador–constructo puede reconfigurarse. Un modelo o interpretación calibrados en el régimen anterior pueden fallar en el nuevo aunque el indicador permanezca "bien calculado".

Juntos, estos mecanismos explican el patrón más importante observado: la estabilidad numérica o la consistencia del modelo dentro de un periodo no garantizan interpretabilidad longitudinal. Lo que se rompe no es el dato; es el puente semántico.

El caso "no presentación": de variable moralizada a variable estratégica

El no presentación/ausentismo estratégico es el punto donde la institución más rápido se equivoca por un sesgo clásico: convertir una decisión situada en un juicio moral. En transiciones tecnológicas, "no rendir" puede expresar gestión de riesgo, reoptimización de tiempos o ajuste al nuevo costo de preparación, más que apatía. Además, el

no presentación es altamente sensible a reglas (recuperatorios, regularidad, promoción) y, por lo tanto, su interpretación requiere contexto.

Este resultado tiene una implicación inmediata: políticas punitivas o basadas en estigmatización tienden a empeorar el fenómeno, porque activan adaptación defensiva. Si se pretende intervenir, debe hacerse sobre el mecanismo (costos, reglas, señales de preparación, acceso a apoyo académico) y no sobre el síntoma moralizado.

Evaluación institucional: pasar de auditoría numérica a auditoría interpretativa

Si el riesgo central es la ruptura inferencial, la evaluación institucional necesita incorporar un componente explícito de **auditoría interpretativa**. Tres ajustes mínimos se desprenden del marco:

- Bitácora de régimen como insumo obligatorio. Toda lectura longitudinal debería estar acompañada por un registro de cambios relevantes (modalidad, plataforma, reglas, instrumentos). Sin esta bitácora, la institución interpreta tendencias como si el mundo no hubiera cambiado.
- Chequeos relacionales y de heterogeneidad, no solo promedios. La señal más informativa no es una tasa aislada, sino el desacople: cuando relaciones históricas entre indicadores se rompen. Asimismo, la reconfiguración de varianza suele revelar adaptación diferencial y brechas socio-técnicas que los promedios ocultan.
- Etiquetado de confiabilidad del indicador (alta/media/baja confiabilidad). Cuando un indicador muestra señales de inestabilidad semántica, debe limitarse su uso decisional y exigir triangulación. Esto reduce el riesgo de políticas basadas en un solo indicador de desempeño y, por ende, reduce Goodhart.

Este punto es crucial para “calidad”: sin esta capa interpretativa, la institución puede exhibir tableros sofisticados y, sin embargo, operar con una epistemología frágil.

Calidad: redefinición operativa bajo transición tecnológica

La calidad no debería evaluarse únicamente por niveles de rendimiento o tasas agregadas, porque esas métricas pueden volverse reactivas. Bajo transición tecnológica, una definición operativa más robusta es: **calidad como capacidad institucional de gobernar con evidencia interpretada bajo conciencia de régimen**.

Esto implica que la calidad incluye, como propiedad del sistema, la existencia de:

- contratos de medición (qué mide cada indicador y bajo qué supuestos),
- trazabilidad de datos/definiciones,
- mecanismos de auditoría interpretativa y triangulación,
- separación entre métricas de diagnóstico y métricas de incentivo.

En otras palabras: no solo importa “qué números dan”, sino si el sistema es capaz de saber cuándo esos números dejaron de significar lo que creía.

Política universitaria: incentivos, equidad y riesgo de daño distributivo

En educación superior, las métricas se convierten rápidamente en política, aun cuando se presenten como “técnicas”. Cuando el significado del indicador es inestable, el riesgo principal no es el error académico: es el **daño distributivo**. Tres implicaciones se derivan de los resultados:

1. Evitar políticas basadas en moralización del rendimiento. Si el sistema cambió el espacio de estrategias, sancionar conductas sin comprender el mecanismo puede incrementar abandono y desigualdad.

2. Reconocer la brecha socio-técnica como componente estructural. La transición tecnológica redistribuye ventajas: acceso, alfabetización digital, tiempo, conectividad. Si la institución interpreta divergencias como "falla individual", interviene tarde y de forma regresiva.

3. Separar métricas para incentivo de métricas para diagnóstico. Una institución que incentiva con un indicador de desempeño dominante acelera Goodhart. Políticas basadas en un conjunto pequeño de indicadores complementarios, con auditoría interpretativa, reducen distorsión y preservan capacidad de medición.

Estas implicaciones son especialmente relevantes para acreditación y rendición de cuentas: la comparabilidad inter-temporal debe tratarse como hipótesis (dependiente del régimen), no como derecho natural del indicador.

Síntesis operativa: reglas de decisión bajo incertidumbre semántica

A partir de los mecanismos y de las firmas empíricas, se derivan reglas prácticas que pueden integrarse a la gobernanza institucional:

- Regla 1: si hay quiebre de régimen, se suspende la atribución automática. Antes de afirmar declive o mejora, se revisa bitácora y se ejecutan chequeos relacionales.
- Regla 2: un indicador sin satélites es un indicador peligroso. Todo indicador crítico debe tener indicadores satélite cuya relación histórica sea monitoreada.
- Regla 3: decisiones punitivas requieren evidencia triangulada. Si el indicador está "amarillo/rojo", se prohíbe usarlo como base única para sanciones o reformas abruptas.
- Regla 4: mirar heterogeneidad antes que promedios. En sistemas adaptativos, la estructura de varianza es información, no ruido.

Estas reglas no requieren grandes presupuestos: requieren disciplina institucional y un estándar epistemológico más alto. La proliferación reciente de políticas universitarias sobre IA generativa refuerza esta exigencia: la institución no sólo debe producir métricas, sino también gobernar sus condiciones de uso e interpretación (Wang et al., 2024).

Arquitectura mínima de medición resiliente: propuesta aplicada

Propósito y alcance: de "tablero" a "sistema de medición"

La discusión previa conduce a una consecuencia práctica: si la transición tecnológica puede producir **deriva semántica** e **inestabilidad inferencial**, entonces una universidad que toma decisiones con métricas necesita algo más robusto que un tablero de indicadores. Necesita una **arquitectura mínima de medición resiliente**: un conjunto acotado de componentes, rutinas y controles que reduzcan el riesgo de interpretar con categorías del régimen anterior.

La propuesta aquí no intenta construir una plataforma completa ni prescribir un modelo único. Su objetivo es más austero (y por eso más útil): definir el **mínimo viable** para que una institución pueda (i) detectar cambios de régimen, (ii) registrar condiciones de contorno relevantes, (iii) auditar la estabilidad de significado de sus métricas, y (iv) sostener decisiones con trazabilidad de supuestos.

Principios de diseño (los "axiomas" de una medición resiliente)

La arquitectura se apoya en cinco principios:

1. Conciencia de régimen: toda inferencia debe explicitar en qué condiciones fue validada y qué cambios podrían invalidarla (Lucas, 1976).
2. Separación entre validez técnica y validez sustantiva: buen ajuste estadístico no implica significado estable; se requieren chequeos semánticos (Meredith, 1993).
3. Auditoría interpretativa, no solo auditoría numérica: monitorear relaciones entre indicadores y supuestos de interpretación (Espeland & Stevens, 2008; Power, 1997).
4. Diseño anti-Goodhart: minimizar incentivos perversos evitando un único indicador "rey" y privilegiando conjuntos pequeños de señales complementarias (Campbell, 1979; Goodhart, 1975).
5. Trazabilidad y reversibilidad: las decisiones basadas en métricas deben dejar un rastro reproducible de datos, supuestos, versiones y contexto; y permitir revisión posterior.

Componentes mínimos de la arquitectura (qué debe existir sí o sí)

La Figura 1 resume los componentes mínimos y el flujo de decisión de esta arquitectura.

Figura 1. Arquitectura mínima de medición resiliente y flujo básico de decisión.

Contrato de medición	**Bitácora de régimen**	**Datos versionados**
↓	↓	↓
Checks semánticos	**Triangulación mínima**	**Gobernanza interpretativa**
↓	↓	↓
Clasificación (verde / amarillo / rojo)		
→ decisión con trazabilidad y confiabilidad		

Fuente: elaboración propia.

Nota. La arquitectura combina contrato de medición, registro de cambios de régimen, trazabilidad de datos, chequeos semánticos y triangulación para limitar inferencias inválidas bajo transición tecnológica.

Contrato de medición

Documento breve, versionado, que define para cada indicador clave:

- Constructo: qué intenta representar (p. ej., progreso, dominio conceptual, persistencia, carga).
- Definición operacional: fórmula exacta, población, periodo, reglas de cómputo.
- Supuestos de invariancia: qué se asume estable en el tiempo (reglas de evaluación, modalidad, acceso, tecnologías).
- Usos permitidos y no permitidos: qué decisiones puede respaldar y cuáles serían extrapolación ilegítima.
- Riesgos de gaming y endogeneidad: cómo puede degradarse bajo presión (Goodhart/Campbell).
- Indicadores satélite: métricas que deberían moverse "coherentemente" con el indicador si conserva significado.

Este contrato reduce el problema clásico de las métricas universitarias: definiciones flotantes y usos expandibles sin control.

Bitácora de régimen

Registro sistemático, con fechas y alcance, de cualquier cambio que altere condiciones de contorno:

- cambios de LMS/entornos digitales, conectividad, autenticación, IA institucional;
- cambios reglamentarios (regularidad, promociones, recuperatorios, correlatividades);
- cambios en modalidad (presencial/híbrida/online), calendarios, turnos;
- cambios de instrumentos de evaluación (tipo de examen, banco de ítems, oral/escrito, antifraude);
- cambios organizacionales relevantes (políticas de retención, tutorías, alertas).

Flujo de procesamiento reproducible y "versionado" de datos/indicadores

Para que la medición sea auditable (no solo presentable), la institución debe sostener:

- versionado de conjuntos de datos (fecha de corte, fuente, transformaciones);
- versionado de definiciones (cambios en fórmulas o limpieza);
- registro de calidad (faltantes, duplicados, cambios en codificación);
- trazabilidad de reportes (qué tablero usó qué versión de datos).

Esto no es lujo técnico: es lo mínimo para que una discusión sobre tendencias no termine en "creo que…".

Chequeos de estabilidad semántica

Conjunto pequeño de pruebas rutinarias, automáticas o semi-automáticas, orientadas a detectar desacoples típicos. Tres familias mínimas:

(a) Chequeos relacionales

Monitorean si se rompen correlaciones o elasticidades históricas entre indicadores. Ejemplos:

- actividad/participación ↔ aprobación;
- parciales ↔ finales;
- regularidad ↔ tiempo a rendir;
- asistencia ↔ desempeño (si aplica).

(b) Chequeos de distribución y heterogeneidad

No mirar solo medias: monitorear varianza, colas, y divergencia entre cohortes/subgrupos (Davis & Sumara, 2006).

(c) Chequeos de cambio de régimen

Detección de cambios de nivel y quiebres de tendencia cercanos a hitos de la bitácora. No como "prueba de causa", sino como alerta semántica.

Triangulación mínima de constructos críticos

Para constructos de alto impacto (abandono, progreso, "rendimiento"), se requiere al menos una triangulación acotada:

- combinar un indicador administrativo (p. ej., aprobación) con otro de distinta exposición a gaming (p. ej., desempeño en un instrumento común interno, o consistencia inter-asignaturas);

- cuando sea posible, incorporar una fuente independiente (p. ej., cargas horarias auto-reportadas con validación, o métricas de trayectoria robustas).

La triangulación no busca "capturar la verdad total"; busca reducir la probabilidad de inferencias semánticamente erróneas.

Comité o rutina de gobernanza interpretativa

No basta con analistas. Se necesita un **mecanismo institucional** de revisión:

- periodicidad (mensual/trimestral);
- roles (datos, académicos, gestión);
- protocolo: revisar bitácora → revisar checks semánticos → revisar implicaciones decisionales → documentar conclusión.

Esto encarna la idea de "audit society" pero en versión saludable: auditoría del significado, no solo del número (Power, 1997).

Flujo operativo recomendado (ciclo de decisión resiliente)

La arquitectura se operacionaliza en un ciclo simple:

1. Cálculo: generar indicadores con flujo de procesamiento versionado.
2. Contextualización: consultar bitácora de régimen para el periodo.
3. Chequeo semántico: ejecutar pruebas mínimas (relacionales, distribución, breaks).
4. Clasificación de confiabilidad: etiquetar cada indicador (verde/amarillo/rojo) según estabilidad.
5. Decisión con trazabilidad: si es verde, se usa con confianza; si es amarillo, se usa con cautela y triangulación; si es rojo, se evita como base única y se prioriza análisis cualitativo/mixto.
6. Registro: documentar supuestos, versión de datos, conclusión y decisión.

El punto clave es introducir una salida explícita que hoy suele faltar: "este indicador está en rojo por cambio de régimen; no se usa para atribuir declive".

Diseño anti-Goodhart: cómo reducir presión destructiva sobre métricas

Para minimizar la degradación por Goodhart/Campbell (Campbell, 1979; Goodhart, 1975), la arquitectura recomienda:

- **No** gobernar con un único indicador de desempeño dominante.
- Usar **pequeños conjuntos** (3–7 indicadores) con roles complementarios: progreso, riesgo, equidad, carga, consistencia.
- Separar métricas para **diagnóstico** de métricas para **incentivo**: lo que se usa para premiar/sancionar debe ser distinto de lo que se usa para comprender.
- Introducir **rotación o revisión** anual de indicadores, con evaluación de efectos secundarios.

Indicadores propuestos (ejemplo de "conjunto mínimo" por función)

Sin pretensión de universalidad, un conjunto mínimo puede organizarse por función:

- Progreso: tasa de avance crediticio por cohorte; ritmo de aprobación en tramos del plan.
- Persistencia: retención año a año; reingreso tras abandono temporal.
- Riesgo evaluativo: no presentación/ausentismo a examen; repetición; “cuello de botella” por asignatura.
- Equidad/heterogeneidad: varianza de ritmo de avance; brechas entre subgrupos.
- Consistencia: desacoples entre indicadores satélite (checks relacionales).

La clave no es la lista, sino que cada indicador tenga contrato, bitácora y chequeo semántico.

Limitaciones y condiciones de implementación

Esta arquitectura mínima no elimina incertidumbre ni reemplaza juicio académico. Sus límites:

- Requiere disciplina organizacional: registrar cambios y versionar datos.
- No resuelve por sí misma disputas normativas sobre qué es “calidad”.
- Necesita adaptación local: el contrato de medición depende del régimen institucional y del plan de estudios.

Aun así, su valor es fuerte: reduce la probabilidad de que una institución confunda cambios de régimen con declive, y evita decisiones de alto impacto basadas en métricas semánticamente inestables.

Cierre del apartado: una práctica institucional de humildad inferencial

La arquitectura propuesta institucionaliza una forma de **humildad inferencial**: la inferencia técnica es necesaria, pero sin auditoría del significado puede ser insuficiente. En tiempos de transición tecnológica, medir no es solo calcular; es sostener un sistema de medición que reconozca que el objeto medido aprende y cambia. La alternativa —gobernar con tableros sin conciencia de régimen— incrementa la probabilidad de diagnósticos erróneos con apariencia de precisión.

Conclusiones, limitaciones del estudio y prospectiva

Conclusiones: lo que este trabajo sostiene (y lo que no)

Este manuscrito partió de una paradoja frecuente en educación superior bajo transición tecnológica: series longitudinales estadísticamente consistentes pueden inducir interpretaciones sustantivas erróneas. El argumento central no es que “las métricas no sirvan”, ni que “todo sea relativo”, sino algo más preciso (y más incómodo): en sistemas universitarios entendidos como **sistemas adaptativos complejos**, la medición es parte del sistema; por lo tanto, cuando cambian las condiciones de contorno (plataformas, modalidades, IA, reglas y costos), puede cambiar el significado operativo de los indicadores aunque su cálculo permanezca intacto (Davis & Sumara, 2006; Espeland & Stevens, 2008).

A partir de la literatura sobre cuantificación y auditoría, se sostuvo que los indicadores no solo describen: **gobiernan**. Al introducir comparabilidad y presión decisional, producen performatividad y reactividad, abriendo la puerta a la dinámica Goodhart/Campbell: cuando una medida se convierte en objetivo, se degrada como medida (Campbell, 1979; Goodhart, 1975; Power, 1997). Bajo transición tecnológica, esta dinámica se intensifica por reducción de costos de adaptación y ampliación del repertorio estratégico.

Los “resultados” se presentaron como un conjunto de **firmas observables** compatibles con cambio de régimen: cambios de nivel, quiebres de tendencia, incremento de heterogeneidad, reordenamiento del no presentación/ausentismo, y —como señal transversal más informativa— **desacoples entre indicadores**. La tesis

discutida fue que estas firmas son advertencias de posible **deriva semántica**, donde un indicador mantiene estabilidad técnica pero pierde validez sustantiva.

Sobre esa base, el aporte aplicado del manuscrito fue proponer una **arquitectura mínima de medición resiliente** que permite a una institución introducir "conciencia de régimen" sin requerir mega-proyectos tecnológicos: contrato de medición, bitácora de régimen, flujo de procesamiento versionado, chequeos de estabilidad semántica, triangulación mínima y gobernanza interpretativa. El objetivo de esta arquitectura es reducir el riesgo de políticas y evaluaciones basadas en interpretaciones rígidas del régimen anterior y, con ello, mitigar daño distributivo y erosión de confianza.

En síntesis, el trabajo concluye que:

1. La inferencia correcta es necesaria pero no suficiente cuando el indicador puede cambiar de significado bajo transición tecnológica.
2. El problema principal no es solo estadístico, sino socio-técnico: endogeneidad, reflexividad y cambio de régimen.
3. La calidad institucional debe incluir la madurez del sistema de medición (capacidad de auditar significado), no solo valores agregados.

Limitaciones del estudio

Este estudio presenta limitaciones derivadas de su diseño conceptual y del objetivo propuesto.

(i) No estimación causal ni cuantificación de magnitudes. El manuscrito no estima el tamaño de la deriva semántica ni identifica causalidad única entre transición tecnológica e indicadores. Propone un marco interpretativo y controles mínimos, más que un modelo causal. En consecuencia, su contribución es principalmente conceptual y aplicada (arquitectura), no inferencial en sentido econométrico.

(ii) Evidencia secundaria y heterogeneidad de contextos. Al basarse en literatura secundaria y en patrones reportados en múltiples contextos, la generalización es teórica (por mecanismo) y no estadística. Las firmas identificadas son plausibles y recurrentes, pero requieren validación localizada para cada institución, carrera o periodo (Espeland & Stevens, 2008; Power, 1997).

(iii) Problema de medición de la propia "deriva semántica". Detectar deriva semántica es metodológicamente desafiante: no se observa directamente, sino por señales indirectas (desacoples, quiebres, variación de relaciones). En ausencia de instrumentos independientes o de diseños cuasi-experimentales, la atribución puede ser ambigua (Meredith, 1993).

(iv) Riesgo de sobrerreacción institucional. Un peligro simétrico existe: usar "cambio de régimen" como excusa para deslegitimar toda evidencia longitudinal. Este manuscrito no propone relativismo; propone **criterios de precaución**. La arquitectura mínima busca precisamente evitar el péndulo: ni confianza ciega ni escepticismo total.

Prospectiva: agenda de investigación y desarrollo institucional

La prospectiva se organiza en dos planos: (a) investigación empírica para validar y cuantificar, y (b) desarrollo institucional para gobernar mejor.

Agenda de investigación empírica (validar y medir la ruptura inferencial)

1. Estudios de invariancia de indicadores administrativos por periodos (pre/post). Adaptar enfoques de invariancia al universo administrativo: analizar estabilidad de relaciones, elasticidades y estructura de varianza por régimen (Meredith, 1993).

2. Diseños cuasi-experimentales donde existan cambios exógenos claros. Aprovechar transiciones tecnológicas con fechas definidas (cambio de LMS, políticas de evaluación online) para explorar discontinuidades y sus mecanismos, evitando atribuciones simplistas.

3. Modelos de trayectorias y heterogeneidad (mixtos y de redes). En lugar de promedios, modelar trayectorias, subpoblaciones y transiciones entre estados académicos. En sistemas adaptativos, la heterogeneidad contiene información causal relevante (Davis & Sumara, 2006).

4. Medición de sustitución de esfuerzo y cambio en costos. Diseñar proxies y mediciones para estimar cuánto del cambio en indicadores se explica por cambios en costos de acceso, automatización y reorganización temporal del estudio.

5. Auditorías de Goodhart en educación superior. Estudiar explícitamente cómo metas institucionales alteran indicadores y qué efectos secundarios generan (Campbell, 1979; Goodhart, 1975). Esto permitiría pasar de advertencias generales a "modos de fallo" específicos por institución.

Agenda institucional (construir capacidad de gobierno bajo transición)

1. Estándar mínimo de medición resiliente. Adoptar formalmente contrato de medición, bitácora de régimen y chequeos semánticos como requisito de cualquier tablero de calidad.

2. Separación entre métricas de incentivo y métricas de diagnóstico. Reducir presión destructiva sobre indicadores críticos; diseñar métricas complementarias para evitar Goodhart.

3. Gobernanza interpretativa y trazabilidad. Institucionalizar revisiones periódicas donde se discuta estabilidad de significado antes de tomar decisiones punitivas o reformas estructurales.

4. Capacitación y alfabetización institucional. Formar a equipos de gestión y académicos en lectura de series longitudinales bajo cambio de régimen, evitando extrapolaciones ingenuas (Lucas, 1976).

5. Ética de la medición y protección contra daño distributivo. Incorporar análisis de impacto diferencial por subpoblación como requisito previo de políticas basadas en indicadores.

Cierre: una tesis sobria para un tiempo de cambios acelerados

La universidad atraviesa una aceleración socio-técnica que reconfigura enseñanza, estudio y evaluación. Por eso, en tiempos de transición tecnológica, la pregunta no es sólo qué muestran los indicadores, sino qué significan ahora bajo un régimen distinto. Responderla exige trazabilidad, conciencia de régimen, auditoría interpretativa y humildad inferencial.

Referencias bibliográficas

Biesta, G. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Paradigm.

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-X

Davis, B., & Sumara, D. (2006). *Complexity and education: Inquiries into learning, teaching, and research*. Lawrence Erlbaum Associates.

Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology, 49*(3), 401–436. https://doi.org/10.1017/S0003975609000150

Goodhart, C. A. E. (1975). *Problems of monetary management: The UK experience* (Vol. 1). Reserve Bank of Australia.

Gough, N. (2013). Towards deconstructive nonalignment: A complexivist view of curriculum, teaching and learning. *South African Journal of Higher Education, 27*(5), 1213–1233.

Lucas, R. E., Jr. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy, 1*, 19–46. https://doi.org/10.1016/S0167-2231(76)80003-6

McConvey, K., & Guha, S. (2024). "This is not a data problem": Algorithms and power in public higher education in Canada. En *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)* (Article 16, pp. 1–14). Association for Computing Machinery. https://doi.org/10.1145/3613904.3642451

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/BF02294825

Muller, J. Z. (2018). *The tyranny of metrics*. Princeton University Press.

Osberg, D., & Biesta, G. (2010). The end/s of education: Complexity and the conundrum of the inclusive educational curriculum. *International Journal of Inclusive Education, 14*(6), 593–607. https://doi.org/10.1080/13603110802530684

Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory, 30*(3), 167–186. https://doi.org/10.1177/0735275112457914

Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence, 7*, 100326. https://doi.org/10.1016/j.caeai.2024.100326

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.1337

Fecha de recepción: 07-01-2026

Fecha de aceptación: 08-04-2026