

Encontrar trabajo formal en Colombia ¿Cuestión de habilidades?

Autor:

Sergio Alejandro Roperó Santiago

Tesis presentada como requisito para optar por el título
de Magíster en Economía

Director:

Darwin Cortés Cortés

Maestría en Economía
Facultad de Economía
Universidad del Rosario
Junio de 2022

Encontrar trabajo formal en Colombia ¿Cuestión de habilidades?¹

Resumen

A través de información del Servicio Público de Empleo (2016-2018) y planes de estudio de los programas de educación superior en Colombia, este trabajo analiza la brecha de habilidades existente entre la oferta educativa y la demanda laboral, enfatizando dicha brecha como una variable relevante para explicar la mayor o menor probabilidad que tienen los egresados de la educación superior para vincularse al mercado laboral formal colombiano. Se presentan los planes de estudio como una fuente de información novedosa para conocer las habilidades ofertadas en el sistema educativo. Además, explotando la heterogeneidad de los planes curriculares, se presenta un modelo de mínimos cuadrados ordinarios con errores robustos a heterocedasticidad que muestra la relación directa entre la oferta de educación superior (oferta laboral potencial) y la demanda laboral.

Clasificación JEL: E24, I23, J64

Palabras clave: Habilidades, brecha de habilidad, educación superior, demanda laboral.

Abstract

Through information from the Public Employment Service (2016-2018) and study plans of higher education programs in Colombia, this paper analyzes the skills gap between educational supply and labor demand, emphasizing said gap as a relevant variable to explain the greater or lesser probability that higher education graduates are linked to the Colombian formal labor market. Study plans are presented as a source of novel information to know the skills offered in the educational system. In addition, exploiting the heterogeneity of the curricular plans, an ordinary least squares model with errors robust to heteroscedasticity is presented, which shows the direct relationship between the supply of higher education (potential labor supply) and labor demand.

JEL Classification: E24, I23, J64

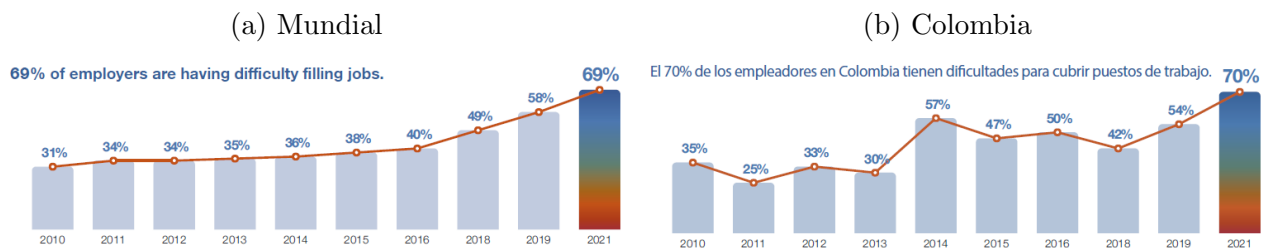
keywords: Skills, skill gap, higher education, labor demand.

¹Agradezco a mi director, Darwin Cortés, por su paciencia, orientación y motivación en el desarrollo de esta investigación. También agradezco a Jorge Gallego, a los participantes del Seminario de Investigación de Economía Informal y del Brown Bag Seminars de la Universidad del Rosario por sus comentarios y aportes a mi investigación. Agradezco a Karol Ríos, Eduardo Rojas, Paula Córdoba y Gabriela Zambrano por su apoyo para la construcción y depuración de la base de planes de estudio con la que fue posible realizar esta investigación. Finalmente, agradezco al Programa Colombia Científica – Alianza EFI por el apoyo financiero del programa inclusión productiva y social: programas y políticas para la promoción de una economía formal, código 60185, que conforma la Alianza EFI – Economía Formal Inclusiva, bajo el Contrato de Recuperación Contingente No FP44842-220-2018. Los aciertos de este documento son sin duda el resultado del intercambio de ideas y opiniones con todos aquellos con quien tuve la oportunidad de conversar sobre mi investigación en los últimos dos años, y cualquier error contenido en este documento es únicamente mi responsabilidad.

1 Introducción

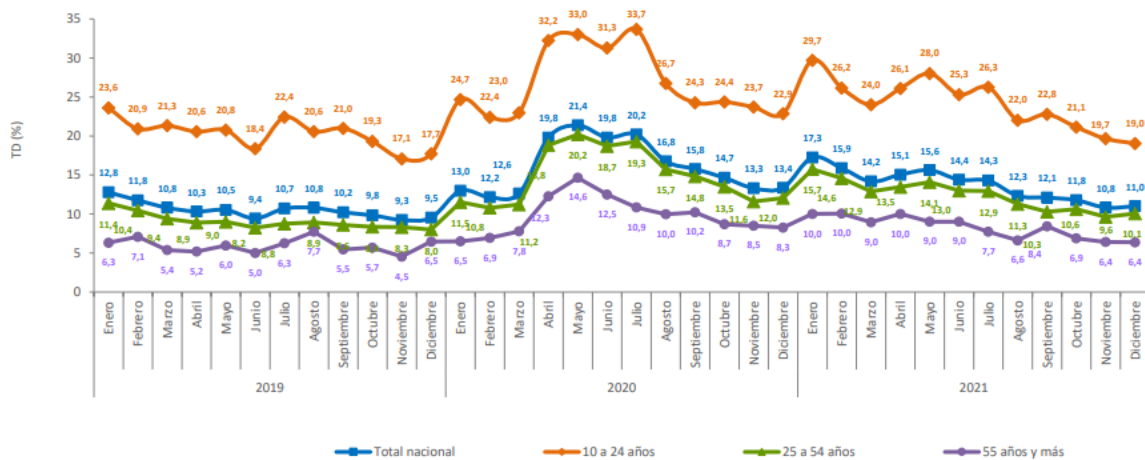
A nivel mundial, la proporción de empleadores que indicaban no encontrar los empleados con las habilidades requeridas era de 30 % en 2009, y en la última década el fenómeno ha venido en aumento (figura 1). Para 2021, el 69 % de los empleadores del mundo indicaron no poder encontrar en el mercado laboral las habilidades que necesitan según el [ManpowerGroup \(2021b\)](#). En Colombia, el [ManpowerGroup \(2021a\)](#) revela una situación similar, 70 % de los empleadores tienen dificultades para cubrir puestos de trabajo. Por otra parte, según las cifras de tasa de desempleo desestacionalizada que publica mensualmente el [DANE \(2021\)](#), el país presenta tasas de desempleo de 2 dígitos desde finales de 2018, alcanzando el máximo histórico de 20,6 % en mayo de 2020 debido a la pandemia de covid-19. También destaca que las personas entre 10 y 24 años registran tasas de desempleo más altas que las de sus contrapartes mayores de 24 años² (Figura 2). Las altas tasas de desempleo que enfrentan los trabajadores y la percepción de los empleadores de no conseguir empleados con las habilidades idóneas para sus vacantes lleva a preguntarse qué está ocurriendo en el sistema educativo donde deberían formarse estas habilidades o, al menos, sus bases.

Figura 1: Dificultad para encontrar trabajadores idóneos



Fuente: Manpower Group - Outlook Survey

Figura 2: Tasa de desempleo por rangos de edad



Fuente: DANE - GEIH marco 2005

²Para mayor información consultar las tasas de desempleo según rangos de edad publicadas por el DANE con base en la GEIH marco 2005

Este trabajo estudia la relación entre la formación de habilidades en el sistema de educación superior, la demanda de habilidades del mercado laboral, y la vinculación laboral formal en Colombia. Específicamente se pregunta ¿Cómo afecta la brecha de habilidades entre la demanda laboral y la oferta de educativa a la tasa de vinculación laboral formal de los egresados de la educación superior? Siendo relevante entender esta relación ya que, los conocimientos³ son la base para desarrollar las habilidades⁴ y, por tanto, contar con un sistema de educación superior alineado con las necesidades del mercado laboral es fundamental para reducir la escasez de habilidades y la tasa de desempleo en el país.

Esta investigación se relaciona con la literatura económica de habilidades iniciada por el trabajo seminal de [Mortensen \(1970\)](#) sobre búsqueda de empleo y las habilidades como fundamento para la determinación del salario. Desde entonces, la literatura económica se ha interesado en estudiar las habilidades desde dos perspectivas: la primera analiza la demanda de habilidades en los mercados laborales e intenta comprender cómo y por qué cambian las habilidades demandadas a través del tiempo. En esta línea de investigación se encuentran los trabajos de [David et al. \(2006\)](#), [Acemoglu and Autor \(2011\)](#), [Hershbein and Kahn \(2018\)](#), [Modestino et al. \(2016\)](#), [Atalay et al. \(2017\)](#) que analizan la demanda de habilidades en economías desarrolladas y encuentran que el progreso tecnológico impulsa un cambio en las habilidades demandadas, cambio que genera un efecto de polarización de la economía, es decir, un proceso donde cada vez más se encuentran cargos de 2 tipos: cargos altos o directivos (trabajos de cuello blanco) que requieren grandes conjuntos de habilidades, o cargos bajos o manuales (trabajos de cuello azul) que requieren pocas habilidades, mientras tienden a desaparecer los cargos medios. Un análisis de la hipótesis de polarización de la economía para el caso colombiano es el trabajo de [Medina and Posso \(2018\)](#).

La segunda línea de estudios sobre habilidades se enfocan en las habilidades socio-emocionales y cómo estas son un buen predictor de las decisiones de escolaridad. En esta literatura se enmarcan los trabajos de [Arcidiacono \(2004\)](#), [Heckman et al. \(2006\)](#), [Carneiro et al. \(2007\)](#), [Almond and Currie \(2011\)](#) y [Todd and Zhang \(2020\)](#). Finalmente, se encuentra el trabajo de [Deming and Noray \(2020\)](#) que estudia cómo el cambio de las habilidades laborales afectan los ingresos profesionales para graduados universitarios y, encuentran que, las carreras de ciencia, tecnología, matemática e ingeniería (STEM por sus siglas en inglés), enseñan habilidades que ofrecen a sus estudiantes una transición relativamente sencilla al mercado laboral, pero que rápidamente se tornan obsoletas ante el cambio tecnológico. Así, existe un ordenamiento de las profesiones elegidas por los estudiantes con base al trade-off entre habilidades altamente demandadas con altas tasas de obsolescencia que son elegidas solo por aquellos con altas capacidades y habilidades menos demandadas pero que se acumulan con la experiencia que son elegidas por aquellos con menores niveles de capacidad.

Ahora, en esta literatura sobre habilidades y sistemas educativos, exceptuando trabajos como el de [Deming and Noray \(2020\)](#), hay poca evidencia que vincule directamente la demanda de habilidades en el mercado laboral con el capital humano aprendido en el sistema educativo. Y no existe, hasta donde el autor conoce, literatura sobre habilidades y sistemas educativos que tengan

³La Clasificación europea de capacidades/competencias, cualificaciones y ocupaciones (ESCO por sus siglas en inglés) entiende por conocimiento el resultado de la asimilación de información a través del aprendizaje.

⁴La ESCO entiende por habilidad la capacidad de aplicar el conocimiento y usarlo para completar tareas y resolver problemas.

en cuenta la heterogeneidad curricular en los sistemas de educación superior. Por tanto, esta investigación aporta evidencia directa de la relación entre lo aprendido en las instituciones de educación superior y la demanda del mercado laboral, considerando la heterogeneidad curricular como un factor fundamental para explicar la mayor o menor facilidad de los graduados de educación superior en el tránsito hacia el mercado laboral.

Esta investigación se enmarca en la teoría de búsqueda y emparejamiento del mercado laboral, que enfatizan la existencia de desequilibrios en el mercado laboral debido a la dificultad de realizar buenos emparejamientos por los problemas de información a los que los agentes se enfrenta y, dadas estas dificultades, las decisiones resultantes pueden ser subóptimas (existencia de malos emparejamientos o de altas tasas de desempleo). Los modelos de búsqueda y emparejamiento enfatizan 3 dificultades de los mercados laborales: 1) los trabajadores forman habilidades sin conocer cuáles serán demandadas por el mercado a futuro, basados, por ejemplo, en su experiencia previa en la escuela (¿qué asignaturas le gustaban en la escuela? ¿en qué área soy bueno?), 2) las empresas demandan habilidades sin saber si estas se encuentren disponibles en el mercado del que participan⁵ y, 3) una vez que trabajadores y firmas han elegido sus conocimiento y habilidades se requiere esfuerzo, representado principalmente en tiempo dedicado a la búsqueda, para encontrar a la contraparte adecuada en el mercado laboral.

Según [Mortensen \(1970\)](#), las habilidades afectan la probabilidad de encontrar trabajo. Esta investigación cuantifica la probabilidad de encontrar trabajo acorde con las habilidades ofrecidas y demandadas en el mercado laboral usando datos para Colombia de las vacantes del servicio público de empleo entre 2016 y 2018 (descripciones de vacantes y códigos CIUO a 4 dígitos⁶), información de los planes de estudio vigentes de las carreras técnicas, tecnológicas y profesionales en Colombia⁷ (2020-2021), características de las IES que ofrecen los planes de estudio (sector, nivel de formación, modalidad, calidad, municipio donde se ofrece el programa, etc.), e información de la tasa de vinculación laboral de recién graduados del Observatorio Laboral para la Educación (OLE) del ministerio de educación nacional para los años 2016 a 2018. Específicamente, se estima un modelo de mínimos cuadrados ordinarios robusto a heterocedasticidad para analizar la relación entre la tasa de vinculación laboral y la brecha de habilidades entre la oferta de educación superior y la demanda laboral.

El reto de esta investigación consistió en estimar la brecha de habilidades. Para superarlo, se empleó la metodología usada por [Sepulveda et al. \(2020\)](#), donde estudian la demanda de habilidades en el mercado laboral de cinco países latinoamericanos a través de minería de texto de descripciones de vacantes de portales web, usando como diccionario de habilidades el listado de 13.485 habilidades de la clasificación europea de capacidades, competencias, cualificaciones y ocupaciones (ESCO, por sus siglas en inglés). Cabe resaltar que, si bien la presenta investigación está relacionada con la de [Sepulveda et al. \(2020\)](#) por la metodología, se diferencia de esta en

⁵[Desjardins and Rubenson \(2011\)](#) encuentran que la demanda laboral es más importante que la oferta a la hora de explicar diferencias en los ingresos, es decir, ofrecer ciertas habilidades en el mercado laboral solo será remunerado a través del salario si el empleo requiere de dichas habilidades

⁶La Clasificación Internacional Uniforme de Ocupaciones (CIUO) cataloga las ocupaciones en grupos temáticos definidos en función de las actividades o tareas principales de cada empleo.

⁷Por plan de estudio vigente se entiende el plan vigente al momento de realizar la consulta y descarga del mismo, que para efectos de esta investigación comprende el periodo entre junio de 2020 y junio de 2021.

varios aspectos, entre ellos: el uso de fuentes de información diferentes⁸, para explicar diferentes variables dependientes⁹ y, el hecho de que esta investigación no solo analiza la demanda sino también la oferta de habilidades.

Después de usar la metodología de diccionario para detectar las habilidades ofrecidas y demandadas, se usó la clasificación CIUO para emparejar ambos lados del mercado laboral, de forma que, se comparó la oferta de cada programa solamente con las ocupaciones que requieren el tipo de conocimientos y/o habilidades que ofrece dicho programa. Una vez unidos ambos lados del mercado laboral se obtuvo, mediante minería de texto, la similitud entre la oferta y la demanda, y la brecha de habilidades se definió como el complemento de dicha medida de similitud. Cabe señalar que esta forma de definir y calcular la brecha de habilidades presenta ventajas y desventajas, dentro de las ventajas se encuentran la disponibilidad de la clasificación ESCO en español para que sirva de diccionario de habilidades, y la facilidad para interpretar la medida (mayor brecha, mayor diferencia entre la oferta y la demanda de habilidades). En contraste, la desventaja que presenta esta metodología es la dificultad para hacer match entre el texto analizado y el diccionario de habilidades, pues al ser una búsqueda exacta de palabras es posible que no se detecten algunas habilidades debido a que se pueden usar sinónimos en los textos analizados, por tanto, las habilidades detectadas representan un mínimo de las habilidades demandadas en el mercado y ofertadas en el sistema educativo. Adicionalmente, como mencionan Sepulveda et al. (2020), en la detección de habilidades pueden presentarse errores de homonimia¹⁰ que deben ser manualmente detectados y eliminados para evitar errores en la medición de las habilidades.

Una vez depuradas las habilidades y calculada la brecha, se estimó un modelo de mínimos cuadrados ordinarios robustos a heterocedasticidad para analizar si una mayor brecha de habilidades disminuye la tasa de vinculación formal de los egresados o no, controlando por variables relevantes como el sector educativo (público o privado), la modalidad del programa (presencial, a distancia o virtual), la calidad educativa (acreditación de alta calidad o no), entre otros. Dado que la tasa de vinculación formal y la brecha de habilidades son porcentajes entre 0 y 100, los resultados de este modelo pueden interpretarse como el efecto promedio en puntos porcentuales de la brecha de habilidades sobre la probabilidad de un egresado de encontrar trabajo en el mercado formal.

Ahora bien, no existe una base de datos pública sobre los planes de estudio de los programas de pregrado en el país, por tanto, otro aporte de esta investigación es la información novedosa sobre la oferta de habilidades en Colombia. Sin embargo, por la posibilidad de acceder a la información, esta investigación solo contempla el 49,74 % de los programas de pregrado del país, es decir 3692 programas de los 7422 programas de pregrado activos en el país para 2020 según el SNIES del ministerio de educación (anexo 1 y 2). Este es un inicio para esta nueva fuente de información

⁸Sepulveda et al. (2020) usan información de *computrabajo* y *elempleo.com*, mientras esta investigación usa información del *servicio público de empleo* y de planes de estudio de pregrado

⁹Sepulveda et al. (2020) tratan de explicar la mayor o menor probabilidad de que una vacante sea potencialmente para una persona joven o no, mientras esta investigación analiza la mayor o menor probabilidad de que un recién egresado del pregrado se vincule en el mercado laboral formal

¹⁰El error de homonimia consiste en un hallazgo perfecto de un término del diccionario ESCO en la descripción, no obstante, el significado del conocimiento o habilidad ESCO no coincide con el contexto utilizado en la descripción del anuncio (Sepulveda et al., 2020)

sobre el sistema educativo colombiano pero, a futuro, se requiere complementar la base de datos para profundizar en la relación entre demanda de habilidades y sistemas educativos, por ejemplo, ampliando la cobertura de los programas analizados, incluyendo programas de posgrado en la base y haciendo un seguimiento de los programas y sus cambios a través del tiempo.

Además de esta introducción, en la sección 2 se presenta un breve contexto del sistema de educación colombiano. La sección 3 presenta y contextualiza los datos utilizados en esta investigación. La estrategia empírica se encuentran en la sección 4. Los resultados se presentan y analizan en la sección 5 y, en la sección 6, se presentan las principales conclusiones e implicaciones de los hallazgos encontrados.

2 Contexto

El sistema de educación de Colombia está compuesto por 5 etapas: 1) educación inicial¹¹, 2) educación preescolar, 3) educación básica (cinco grados de primaria y cuatro grados de secundaria), 4) educación media (dos grados)¹², y 5) educación superior¹³ (figura 3). En Colombia, la educación superior se divide en 2 niveles: pregrado y posgrado. A su vez, el nivel de pregrado se subdivide en 3 niveles: técnico, tecnológico y profesional. De igual forma, el nivel de posgrado se subdivide en 3 niveles: especializaciones, maestrías y doctorados.

La educación superior a nivel de pregrado es el foco de esta investigación, por tanto, es necesario aclarar algunos aspectos del funcionamiento de dicho nivel educativo. En primer lugar, en Colombia contar con un título de bachiller y haber presentado el examen de estado para el ingreso a la educación superior (Saber 11) son requisitos indispensables para acceder a la educación superior, es decir, solo pueden ingresar a los programas de pregrado aquellas personas que han culminado la educación media. En segundo lugar, los niveles de formación de pregrado son paralelos (no jerárquicos), lo que significa que un bachiller puede acceder a cualquier de los 3 niveles de formación de pregrado. En tercer lugar, el sistema de educación superior colombiano está compuesto por 3 tipos de instituciones: 1) instituciones técnicas profesionales, que pueden ofrecer programas técnicos; 2) instituciones universitarias o escuelas tecnológicas, que pueden ofrecer programas técnicos y tecnológicos; y 3) universidades, que pueden ofrecer programas técnicos, tecnológicos y profesionales.

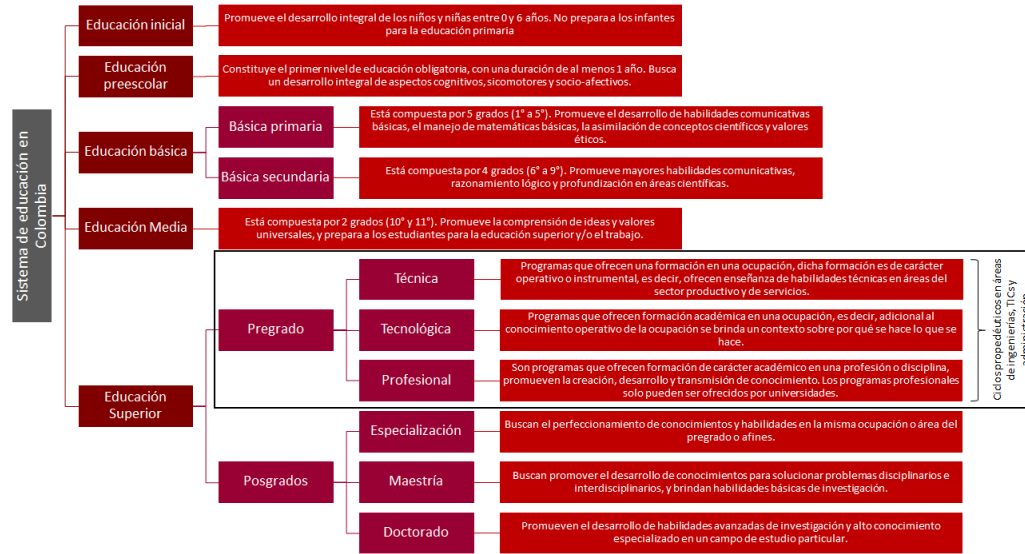
Sin embargo, el sistema colombiano permite que las instituciones técnicas y tecnológicas ofrezcan educación equivalente a la profesional siempre y cuando esta sea ofrecida a través de ciclos propedéuticos, y que dichos ciclos sean derivados de los programas que están habilitadas por la ley para ofrecer. Adicionalmente, los ciclos propedéuticos solo se permiten en las áreas de ingeniería, TIC y administración ([ley 749 de 2002](#)).

¹¹Para más información sobre la educación inicial consultar la página web del [ministerio de educación](#).

¹²Para más información sobre la educación preescolar, básica y media consultar la [ley general de educación \(ley 115 de 1994, artículos 15, 19 y 27\)](#).

¹³Para más información sobre la educación superior consultar la [ley de educación superior \(ley 30 de 1992, artículos 9 al 13 y 17 al 19\)](#). Adicionalmente, consultar la [ley 749 de 2002](#), por la cual se organiza el servicio público de la educación superior en las modalidades de formación técnica profesional y tecnológica, y se dictan otras disposiciones.

Figura 3: Esquema del sistema de educación colombiano



Nota: Esta gráfica presenta la estructura del sistema de educación colombiano y una breve explicación del objetivo de cada nivel educativo. La educación inicial es el único nivel educativo que no hace parte del sistema de educación formal, pues no prepara a los y las estudiantes para niveles de educación posteriores. Para la construcción del esquema se usó información del ministerio de educación, la ley 30 de 1992, la ley 115 de 1994 y la ley 749 de 2022.

3 Datos

Para analizar cómo la brecha de habilidades afecta a la vinculación laboral de los egresados de pregrado en Colombia se requieren tres insumos: información de la oferta de habilidades, información de la demanda de habilidades y una medida de vinculación laboral de los egresados del sistema de educación superior. La oferta de habilidades se aproximó a través de los nombres de las asignaturas que conforman los planes de estudio de las carreras de pregrado. La demanda de habilidades se obtuvo del análisis de las descripciones de las vacantes del servicio público de empleo. Para ambas fuentes se usó el listado de conocimientos y habilidades ESCO como diccionario para identificar las habilidades ofrecidas y demandadas con las cuales se calculó la brecha de habilidades. Finalmente, para la medida de vinculación laboral se usó la tasa de vinculación laboral del Observatorio Laboral para la Educación del Ministerio de Educación.

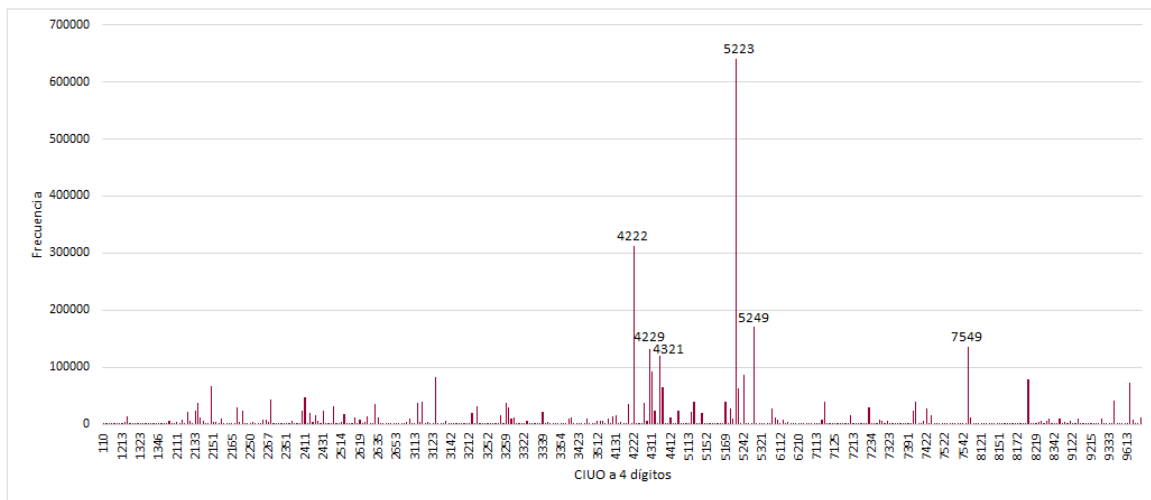
3.1 Vacantes del Servicio Público de Empleo

Al igual que la información de vacantes de portales web que se usan cada vez más en las investigaciones sobre habilidades en el mercado laboral, las vacantes registradas en el Servicio Público de Empleo (SPE) ofrecen información a bajo costo y en periodos de tiempo más cortos de los que puede ofrecer una encuesta o censo. Así, el SPE es fuente de información útil para

analizar tendencias de mercado en periodos cortos de tiempo, y ofrecer a diferentes usuarios datos relevantes, por ejemplo, brindar información depurada de vacantes compatibles con los perfiles de habilidades de quienes buscan empleo, o para informar a los hacedores de política sobre las brechas existentes en el mercado laboral.

En el trabajo de **Sepúlveda et. al. (2020)** se menciona que para que un portal web como el del SPE sea relevante para investigaciones del mercado laboral este debe cumplir 4 criterios: dinamismo del sitio web, cantidad de vacantes publicadas, número de campos relevantes y calidad de la información, entre mejor desempeño presenten estos 4 indicadores, mejor será la fuente de información. **Sepúlveda et. al. (2020)** no usan información del servicio público de empleo debido al poco dinamismo de su portal web. Sin embargo, desde hace algunos años el SPE vincula una red de prestadores de servicio que incluye entidades tradicionales de empleo, tanto públicas como privadas, y portales web de empleo como *computrabajo* o *el empleo.com* lo que hace de esta fuente una de las fuentes con mayor número de vacantes del país. Además, La página web del SPE es el tercer portal de empleo más visitado en Colombia, y las vacantes publicadas allí abarcan un amplio espectro de ocupaciones como puede observarse en la gráfica 4.

Figura 4: Número de anuncios de vacantes según CIUO a 4 dígitos



Nota: Esta gráfica presenta la frecuencia absoluta con la que aparecen las vacantes del Servicio Público de Empleo según su clasificación en la CIUO a 4 dígitos. No se eliminaron anuncios que puedan considerarse duplicados pues las empresas con alta rotación tienden a publicar anuncios iguales y/o similares en diferentes momentos del tiempo, es decir, son vacantes diferentes (ej. CIUO 5223 - Asistentes de venta de tiendas y almacenes).

Es de resaltar que el SPE presenta campos estandarizados para publicar una oferta laboral como son el título de la oferta, la descripción, el salario, el municipio donde se ubica el empleo, los requerimientos de educación y experiencia, entre otros. También debe señalarse que el SPE clasifica las vacantes según la Clasificación Internacional Uniforme de Ocupaciones (CIUO). Por todo lo anterior, en la actualidad las bases de datos de vacantes del SPE son una fuente útil y pertinente para los fines de la presente investigación, en particular porque esta se centra en

el mercado laboral formal. Alrededor del 45 % a 50 % del empleo en las 23 ciudades principales del país es informal según cifras del [DANE \(2021\)](#), en contraste las vacantes publicadas en el SPE cuando mencionan la remuneración en las descripciones señalan aspectos como: “se pagan prestaciones de ley.” “salario + prestaciones”, es decir, tienen un carácter más formal.

Las variables más relevantes para esta investigación son: 1) la descripción de la vacante que, luego de un proceso de eliminación de tildes, mayúsculas y palabras sin información como artículos, preposiciones o conectores, se usa para extraer las habilidades requeridas por las vacantes a través de técnicas de minería de texto; y 2) el código CIUO a 4 dígitos que será usado para cruzar la información con la oferta de habilidades de los planes de estudio de los programas de pregrado.

3.2 Planes de estudio

Los planes curriculares de las carreras técnicas, tecnológicas y profesionales son una fuente novedosa para aproximarse a la oferta de habilidades del sistema educativo, y nos ofrecen la posibilidad de llenar el vacío de información que existe a la hora de estudiar las habilidades de los futuros trabajadores. Los planes de estudio reportan los nombres de las asignaturas que los estudiantes deben cursar como requisito para obtener un título específico y, considerando que los nombres de las asignaturas refieren fundamentalmente a las áreas de estudio que se abordarán en dichas asignaturas, los planes sirven como una aproximación a los conocimientos y habilidades que un estudiante debe adquirir o desarrollar cuando decide estudiar una carrera determinada.

Ahora bien, no existe una base de datos consolidada con los planes de estudio de los programas de pregrado ofrecidos en Colombia. Por tanto, es relevante señalar que la construcción de la base de planes de estudio utilizada en esta investigación partió del registro SNIES del Ministerio de Educación. En este registro se cuenta con información de todos los programas activos que se ofrecían en el país a corte de junio de 2021 y, para cada programa se tiene información de la IES que lo ofrece, nombre del programa, título que otorga, sector al que pertenece (público o privado), nivel de formación del programa (técnico, tecnología, universitario), modalidad de estudio (presencial, distancia tradicional, distancia virtual), si el programa está o no acreditado como de alta calidad y el municipio donde se ofrece el programa, entre otras variables.

Al registro SNIES se agregó una variable adicional, el plan de estudios. Para esto, manualmente se descargaron de las páginas web de las IES los planes de estudio de cada programa¹⁴ diferenciando por nivel de formación (para aquellas carreras ofrecidas por ciclos propedéuticos), modalidad (para aquellos programas que una misma IES ofrece a través de varias modalidades) y municipio (para aquellas IES que tienen sedes en múltiples municipios del país)¹⁵. Una vez consolidado el conjunto de planes de estudio se usó reconocimiento óptico de caracteres (OCR, por sus siglas en inglés) para transformar los archivos de formato PDF o imágenes (JPG o PNG)

¹⁴Los planes de estudio fueron descargados manualmente debido a la heterogeneidad existente en las estructuras de las páginas web de las universidades que no permite utilizar técnicas de web scraping para la descarga automática de información.

¹⁵Ver anexo 2 para detallar mejor las diferencias entre los programas seleccionados en la muestra y aquellos que no fueron incluidos en la muestra

en texto que se agregó a la nueva variable¹⁶.

Sin embargo, como señalan [Atalay et al. \(2017\)](#), el proceso de OCR puede cometer errores al momento de detectar algunas letras, caracteres o palabras (gráficas 5 y 6). Por tanto, una vez construida la variable planes de estudio se requirió una verificación manual para asegurar la calidad del texto que se usó para obtener las habilidades ofrecidas por los programas de pregrado (gráfica 7). De forma análoga a lo que se hizo con las vacantes del SPE, el texto se transformó eliminando tildes, mayúsculas y palabras sin información. Con el texto depurado de esta manera, se usó minería de textos en la variable de planes de estudio para la extracción de las habilidades ofertadas.

Figura 5: Ejemplo de plan de estudios - Biología de la Universidad del Rosario.

I	II	III	IV	V	VI	VII	VIII
Biología de los cambios globales (3)	Evolución de la vida (3)	Botánica (4)	Ecología (4)	Sistemas de información geográfica (3)	Sistemas socio-ecológicos (4)	Profundización I (4)	Profundización III (4)
Física I (3)	Física II (3)	Biología celular (3)	Biología molecular (4)	Genética (4)	Fund. derecho y política ambiental (3)	Profundización II (4)	Opción de grado * (6)
Química general (3)	Química orgánica (3)	Bioquímica funcional (4)	Zoología de vertebrados (3)	Zoología Invertebrados (3)	Evolución (3)	Electiva Profesional (4)	Electiva profesional (4)
Cálculo diferencial (3)	Cálculo integral (3)	Biomatemáticas (3)	Biología de microorganismos (3)	Fisiología Vegetal (4)	Conservación y restauración (3)	Bioestadística (4)	Electiva no profesional (2)
Programación de computadores (3)	Fundamentos de análisis de datos (3)	Electiva no profesional (2)	Constitución política e instrucción cívica (2)	Fisiología animal (3)	Biología del desarrollo (3)	Ética (2)	Electiva no profesional (2)
Cátedra rosarista (2)	Núcleo de formación rosarista (2)	Núcleo de formación rosarista (2)	Núcleo de formación rosarista (2)	Redacción de manuscritos científicos (1)	Constitución política e instrucción cívica (2)		
Módulo integrador I (1)	Módulo integrador II (1)						
Cultura rosarista (0)							

* Existen tres opciones de grado:
Tesis de investigación corta (6 créditos)
Una pasantía en una empresa o una organización (6 créditos)
Una opción coterminal con una maestría (14 créditos de cursos de nivel de posgrado)

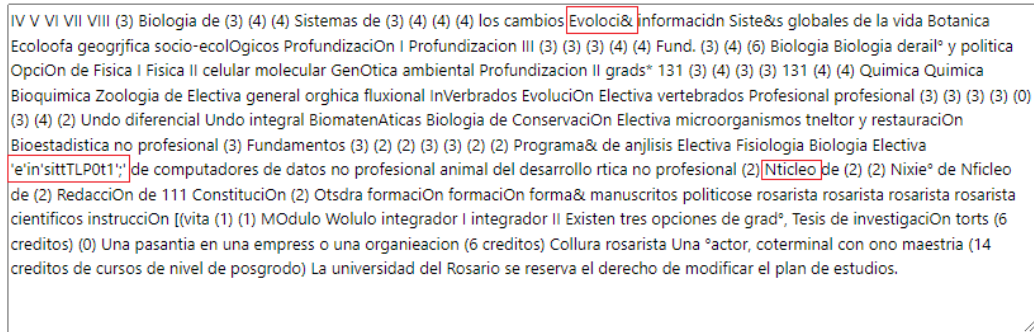
La universidad del Rosario se reserva el derecho de modificar el plan de estudios.

Fuente: Universidad del Rosario

Nota: Los planes de estudio presentan la estructura curricular de los programas de educación superior, es decir, listan las asignaturas que los estudiantes deben cursar a lo largo del programa. Adicionalmente, presentan la forma recomendada en que el estudiante debe cursar dichas asignaturas en los diferentes semestres considerando los prerrequisitos y correquisitos de conocimientos necesarios para cursar efectivamente una asignatura.

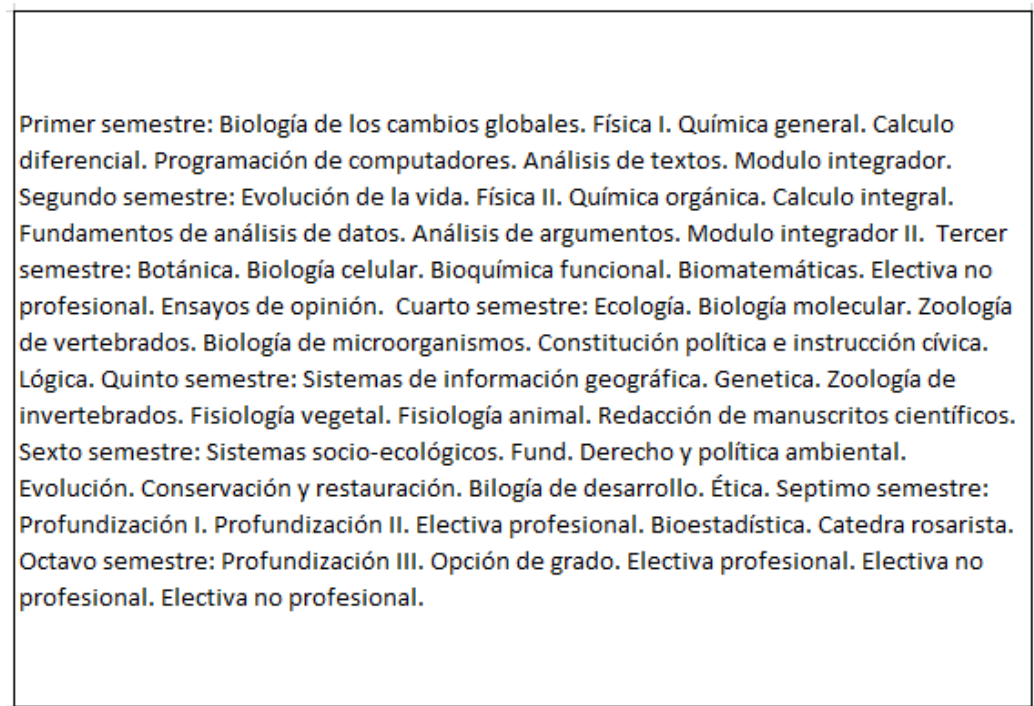
¹⁶Para la aplicación de OCR sobre los planes de estudio se usó el convertidor onlineocr.net

Figura 6: Ejemplo de OCR sobre un plan de estudios - Biología de la Universidad del Rosario.



Nota: Esta gráfica presenta un ejemplo del resultado del reconocimiento óptico de caracteres (OCR) aplicado al plan de estudios del programa de Biología de la Universidad del Rosario. Resaltado en recuadros rojos se muestran algunos errores existentes en el proceso de OCR que deben ser corregidos manualmente antes de continuar con la extracción de habilidades.

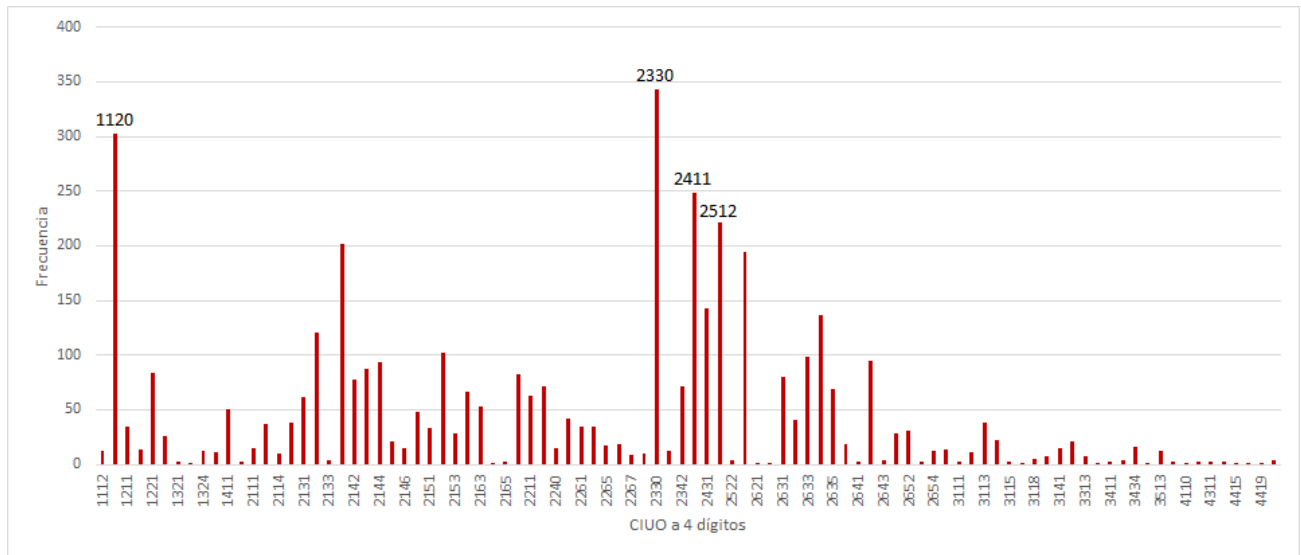
Figura 7: Ejemplo de OCR corregido manualmente - Biología de la Universidad del Rosario.



Nota: Esta gráfica presenta un ejemplo del resultado del reconocimiento óptico de caracteres (OCR) aplicado al plan de estudios del programa de Biología de la Universidad del Rosario tras ser verificado y corregido manualmente. Se eliminaron los números que no aportaban información y se corrigieron todos los errores en la detección de caracteres del OCR.

También se construyó la variable CIUO a 4 dígitos en la base de planes de estudio a partir de los títulos otorgados por los programas ¹⁷. Para hacerlo se buscaron los títulos otorgados en la base de ocupaciones de la ESCO y se asignaron los códigos CIUO que la ESCO otorga a dichas ocupaciones. Para la mayoría de ocupaciones la asignación fue inequívoca dado que la ESCO presenta una etiqueta principal para las ocupaciones y etiquetas alternativas o ejemplo de ocupaciones clasificadas en los CIUO. Sin embargo, para aquellos títulos cuya clasificación presentaba dificultad se usó la descripción de las ocupaciones de la ESCO, el documento de la clasificación única de ocupaciones para Colombia (CUOC) y los perfiles profesionales u ocupacionales de los programas para determinar la mejor clasificación posible.

Figura 8: Número de programas según CIUO a 4 dígitos.



Nota: Esta gráfica presenta la frecuencia absoluta con la que aparecen los planes de estudio recopilados en la muestra de la investigación según su clasificación en la CIUO a 4 dígitos. la gráfica está truncada en el gran grupo de actividades 5, es decir, no presenta ningún plan de estudios de educación superior en los grandes grupos de actividad 6, 7, 8 o 9, esto se debe a que las ocupaciones en los grandes grupos 7 a 9 requieren niveles de habilidades que no necesitan de educación superior para ser adquiridas. La ocupaciones del gran grupo 6 son un punto de bisagra entre las actividades que requieren y las que no requieren educación superior, por tanto, es problemático determinar programas de educación superior que pertenezcan al gran grupo de ocupación 6.

La gráfica 8 muestra el número de programas clasificados según CIUO a 4 dígitos. En ella se puede observar una distribución truncada después del quinto gran grupo de ocupaciones ¹⁸. Sin embargo, al considerar que los grandes grupos 7, 8 y 9 son ocupaciones que requieren niveles

¹⁷De forma análoga a la práctica estándar de usar los títulos de las vacantes para clasificarlas en códigos CIUO a 4 dígitos, se usaron los títulos otorgados por los programas para clasificar los programas en códigos CIUO a 4 dígitos

¹⁸Los grandes grupos de ocupaciones son el nivel más agregado de la CIUO, y se identifican mediante códigos de 1 dígito. Los grandes grupos de ocupación son: (0) Ocupaciones militares, (1) Directores y gerentes, (2) Profesionales científicos e intelectuales, (3) Técnicos y profesionales de nivel medio, (4) Personal de apoyo administrativo, (5) Trabajadores de los servicios y vendedores de comercio y mercados, (6) Agricultores y trabajadores calificados agropecuarios, forestales y pesqueros, (7) Oficiales, operarios y artesanos de artes mecánicas y de otros oficios, (8) Operadores de instalaciones y máquinas y ensambladores, (9) ocupaciones elementales.

de competencia 1 o 2, es decir, trabajos manuales o de operación de maquinas que requieren solo habilidades básicas de lectura y matemáticas es de esperarse que los títulos de educación superior no se clasifiquen en estos códigos, ya que, alcanzar este nivel de competencias no requiere educación técnica, tecnológica o universitaria.

3.3 Diccionario de conocimientos y habilidades

La Clasificación Europea de capacidades/competencias, cualificaciones y ocupaciones (ESCO) “determina y categoriza las capacidades, competencias, cualificaciones y ocupaciones pertinentes para el mercado laboral y la educación y formación en la Unión Europea”. Esta clasificación muestra las relaciones entre las ocupaciones y las capacidades y cualificaciones, en particular asocia 13.485 conocimientos y habilidades a 2.942 ocupaciones. Además de distinguir entre conocimientos y habilidades, la ESCO clasifica cada conocimiento o habilidad según el nivel de reutilización de la misma, es decir, cuán ampliamente se puede aplicar un conocimiento o habilidad.

Para aclarar algunos conceptos, es necesario indicar el significado que la ESCO da a conocimiento, habilidad, y los niveles de reutilización:

- **Conocimiento:** es el resultado de la asimilación de información a través del aprendizaje.
- **Habilidad:** es la capacidad de aplicar conocimientos y utilizar los conocimientos técnicos para completar tareas y resolver problemas.
- **Nivel de reutilización:** cuán ampliamente se puede aplicar un conocimiento o habilidad. Existen 4 niveles para la ESCO.
 - **Transversal:** relevantes para una amplia gama de ocupaciones y sectores económicos (habilidades básicas o blandas).
 - **Intersectorial:** Relevantes para ocupaciones en varios sectores económicos.
 - **Específicas del sector:** Relevantes para más de una ocupación dentro de un mismo sector.
 - **Específicas de la ocupación:** Se aplican solo dentro de una ocupación y sus especialidades.

Además de la utilidad de la ESCO al clasificar las habilidades y conocimientos por nivel de reutilización y relacionarlas con las ocupaciones para el mercado laboral y la educación, el hecho de ser una clasificación basada en la CIUO es fundamental para esta investigación dado que, tanto las vacantes como los planes de estudio fueron clasificados según la CIUO. Por esto, se eligió la ESCO como diccionario de habilidades en lugar de opciones alternativas como la O*NET (cuya clasificación no se basa en la CIUO). También se debe considerar que la ESCO es una clasificación multiligüe por lo que se usó la versión en español para esta investigación.

3.4 Tasa de vinculación laboral

El Observatorio Laboral para la Educación (OLE) del Ministerio de Educación monitorea al tiempo la oferta de graduados en educación superior e indicadores de mercado laboral de fuentes externas (al ministerio). Entre los indicadores externos da seguimiento a la tasa de cotizantes al sistema general de seguridad social (SGSS) y, considerando que en general las personas cotizan solo cuando están trabajando (trabajadores con contratos laborales o independientes), podemos usar la tasa de cotización como una tasa de vinculación laboral formal. Lo anterior es particularmente cierto cuando se analiza la tasa de cotización solo de los recién egresados, pues estos trabajadores que recién ingresan al mercado laboral no tienen razones para cotizar antes de conseguir un trabajo.

Además, la información de la tasa de cotización del OLE se ofrece públicamente desagregada por programa, y puede unirse al registro SNIES mencionado en la sección de planes de estudio a través del código SNIES (que es un identificador único por programa). También se debe señalar que la información del OLE está disponible de 2015 a 2018, por tanto, es posible realizar análisis temporales respecto a la tasa de vinculación.

4 Metodología

En esta sección se profundiza en la construcción de las bases de habilidades demandadas (vacantes) y ofrecidas (planes de estudio) y el cálculo de la brecha de habilidades. Posteriormente se presenta el modelo lineal con el cual se busca responder a la pregunta de investigación planteada.

4.1 Cálculo de la brecha de habilidad

Para calcular la brecha de habilidades se requiere en primera instancia detectar las habilidades ofrecidas y demandadas en el mercado laboral. Como se indicó en subsecciones anteriores esto se hizo a través de minería de texto, usando la ESCO como diccionario. El algoritmo para construir las bases de habilidades se presenta a continuación:

Algoritmo de construcción de bases de brecha de habilidad

1. Se crean 13.485 variables dummies, una por cada habilidad o conocimiento registrado en la ESCO.
2. Se usa minería de texto en el programa Rstudio para buscar las habilidades del diccionario ESCO en la descripción de las vacantes y los planes de estudio. Así, si una habilidad es detectada en la descripción se asigna un 1 en la dummy correspondiente.
3. Se crea una variable colapsada donde se nombran todas las habilidades detectadas para cada una de las vacantes y planes de estudio.
4. Se clasifican las vacantes y los planes de estudio por CIUO a 4 dígitos.

5. Se cruzan las bases de datos a partir de los CIUO. Cada programa se cruza con todas las vacantes clasificadas en el mismo CIUO a 2 dígitos.
6. Se usa minería de texto para buscar las habilidades detectadas en los planes de estudio dentro de las habilidades detectadas en las descripciones de las vacantes. Este proceso es hecho para cada programa respecto a las n-ésimas vacantes con las que fue emparejada.
7. Para cada programa respecto a las n-ésimas vacantes con las que fue emparejada, se contabiliza el número de coincidencias, y se calcula la proporción de similitud.

$$Similitud = \frac{\#Hab.Detectadas.coincidentes}{\#Hab.Requeridas.vacante}$$

8. Para cada programa respecto a las n-ésimas vacantes con las que fue emparejada, se calcula la brecha de habilidad como el complemento de la proporción de similitud.

$$BH = 1 - Similitud$$

9. Se calcula el promedio de brechas existentes entre un programa y las n-ésimas vacantes con las que fue comparada. Es esta forma para cada programa la base resultante solo tiene un dato de brecha (brecha promedio).

El resultado de los primeros 3 pasos del algoritmo es una base de datos estructurada como se muestra en la gráfica 9, donde se cuenta con información individual de cada una de las 13485 habilidades de la ESCO y se construye la variable colapsada de todas las habilidades solicitadas en cada descripción o potencialmente ofrecidas en cada programa. El resultado de los siguientes 5 pasos se presenta en la gráfica 10, siendo la variable “Brecha” la variable de interés que usaremos en el modelo lineal que se presenta en la siguiente sección.

Figura 9: Estructura de las bases de conocimientos y habilidades ESCO.

	Hab. 1	Hab. 2	Hab. 3	...	Hab. 13485	Habilidades
Descripción vacante 1	0	0	1	...	0	Hab. 3
Descripción vacante 2	1	1	0	...	1	Hab. 1, Hab. 2, Hab. 13485
Descripción vacante 3	0	1	0	...	1	Hab. 2, Hab. 13485
⋮	⋮	⋮	⋮	\	⋮	⋮
Descripción vacante n	0	1	1	...	0	Hab. 2, Hab. 3

Nota: Esta gráfica presenta la estructura inicial de la base de datos construida para la extracción de habilidades en la cual se construyeron 13.485 variables con asignación 0 o 1 según la detección o no de la habilidad buscada. La última columna de la matriz de extracción de habilidades resume todas las habilidades detectadas en una vacante separadas con comas. Cabe señalar que existe una matriz de extracción de habilidades para los planes de estudio análoga a la acá ejemplificada para las vacantes.

Figura 10: Cálculo de la brecha de habilidades.

	Habilidades Plan	Habilidades Vacante	Silimitud	Brecha
Plan Estudios 1	Hab. 2	Hab. 3	0	1
Plan Estudios 2	Hab. 1, Hab. 13485	Hab. 1, Hab. 2, Hab. 13485	0,66	0,33
Plan Estudios 3	Hab. 13485	Hab. 2, Hab. 13485	0,5	0,5
⋮	⋮	⋮	⋮	⋮
Plan Estudios n	Hab. 2, Hab. 3	Hab. 2, Hab. 3	1	0

Nota: Esta gráfica presenta la estructura de la base de datos usada para el cálculo de la brecha de habilidades. La base de la brecha de habilidades reúne en una sola matriz los resúmenes de habilidades detectadas en la matriz de extracción de habilidades de vacantes y en la matriz de extracción de habilidades de planes de estudio. A partir de la información de las 2 columnas resumen de habilidades se procede a calcular la similitud y la brecha como se indica en los pasos 7 y 8 del algoritmo de construcción de bases de brecha de habilidades.

Cabe señalar que la detección de habilidades del paso 3 del algoritmo no es perfecta, por tanto, antes de continuar con el proceso de cálculo de la proporción de similitud y, finalmente, el cálculo de la brecha es necesaria una revisión manual de las habilidades detectadas. [Sepulveda et al. \(2020\)](#) señalan que la indexación de habilidades a través de un diccionario como la ESCO puede ocasionar errores de homonimia o de contención. Los errores de homonimia consisten en palabras iguales que presentan acepciones diferentes en la ESCO y en el contexto de las vacantes o planes de estudio, por ejemplo, mientras en la ESCO “equipo de oficina” es un conocimiento sobre la oferta de indumentaria para oficinas, en las vacantes del SPE se encontraban por separado las palabras “equipo” y “oficina” que el algoritmo detectaba como la existencia de una demanda por este conocimiento. Por otra parte, el error más común al detectar habilidades fue el de contención, el algoritmo reconocía en parte de alguna palabra algún conocimiento o habilidad, por ejemplo, el conocimiento en programas estadísticos como R o en el lenguaje de programación C fue casi universalmente detectado en todas las vacantes y planes de estudios porque el algoritmo reconocía cualquier “c” o “r” dentro de una palabra como una demanda u oferta de estos conocimientos. De igual forma, fue reconocido frecuentemente el conocimiento “fines” (idioma) debido a la presencia de palabras como “afines” en las vacantes al referirse que buscaban trabajadores de ciertas profesiones u áreas afines.

Para evitar el error de medición que pueden introducir estos errores de homonimia o contención en el cálculo de la brecha de habilidades se tomó la decisión de eliminarlos de la base de datos. La depuración se hizo analizando el top 20 de habilidades y conocimientos demandados en cada mes de las vacantes del SPE, así se eliminaron los errores más frecuentes. Sin embargo, la ESCO cuenta con una amplia variedad de habilidades correlacionadas, luego, la eliminación de estas habilidades problemáticas no representa un problema para la consistencia de las habilidades detectadas pues otras sin problemas de homonimia o contención capturan conocimientos y habilidades relacionadas con las que fueron eliminadas. Adicionalmente, el hecho de depurar a partir del top 20 de cada uno de los 36 meses de estudio representa una buena limpieza de la base dado que las frecuencias de detección de habilidades cae al alejarse de este top, es decir, por debajo del top 20 de habilidades solo se detectan en 2% o menos de las vacantes totales del mes. Sin embargo, ampliar la depuración hasta el top 50 o 100 permite disminuir aún más

cualquier posible error de medición.

Tabla 1. Habilidades y conocimientos eliminados tras la depuración del top 20.

R	C++	iOS
APL	Finés	Oración
Tala	Galés	id Tech
Estar al día de los productos	Empresa social	gestión del personal
Torá	Equipo de oficina	Lógica
Política	Formación en línea	

Nota: Esta tabla lista las habilidades que fueron eliminadas de las matrices de extracción de habilidades debido a problemas de homonimia o de contención. Por tanto, estas habilidades no fueron tenidas en cuenta en los análisis de brechas de habilidades.

4.2 Estrategia empírica

Se estima un modelo lineal por mínimos cuadrados ordinarios con errores robustos a heterocedasticidad donde la variable dependiente es la tasa de vinculación laboral formal por programa (TVf_p) y la variable independiente de interés es la brecha de habilidad ($BrechaHabilidad_p$) calculada como se explicó en la subsección anterior. Además, se usarán como controles de la regresión las variables $Sector_p$ y $calidad_p$ (medida a través de los puntajes en Saber Pro y Tyt) para controlar las diferencias de calidad educativa y entre la educación pública y privada del país. También se usan como controles el nivel de formación (técnico, tecnólogo o profesional), la modalidad (presencial, distancia o virtual), los grandes grupos de ocupaciones (CIUO a 1 dígito), el área del conocimiento al que pertenece el programa, el número de egresados de un programa para controlar el efecto red/competencia generado por los pares profesionales y efectos fijos de municipio y de año, que serán representados por la matriz \mathbf{X} . La ecuación a estimar es:

$$TVf_p = \beta_0 + \beta_1 BrechaHabilidad_p + \beta_2 sector_p + \beta_3 calidad_p + \beta X + \varepsilon$$

La hipótesis previa a la estimación de resultados es que: a mayor brecha de habilidad entre un programa y la demanda laboral de las ocupaciones relacionadas con este, menor será la tasa de vinculación laboral que tendrán sus egresados. Sin embargo, se espera que estos resultados sean sensibles a características de los programas educativos como la calidad del programa, bastante heterogénea en el sistema de educación colombiano. También se espera cierta sensibilidad al sector de la economía relacionado con el programa (CIUO) pues [Deming and Noray \(2020\)](#) señalan que, ocupaciones del sector STEAM favorecen el tránsito hacia el mercado laboral. Finalmente, se espera resultados diferenciados por municipio donde se ofrece el programa ya que, la vinculación laboral depende del tamaño del mercado laboral del municipio donde se ofrece el programa, aunque este efecto puede ser reducido siempre que exista desplazamiento laboral entre los municipios del país.

5 Resultados

5.1 Oferta y Demanda de habilidades

El objetivo de esta investigación es comprender la mayor o menor probabilidad de que los egresado se vinculen al mercado laboral formal según la brecha que existen entre lo que aprenden en la educación superior y lo que demanda el mercado laboral. Por tanto, el primer paso es conocer qué demanda el mercado laboral y qué ofrece el sistema de educación superior colombiano. La tabla 2 resume el número promedio de cada tipo de habilidad y conocimiento que se demanda y ofrece en Colombia, se observa, por ejemplo, que la mayor demanda del mercado laboral es por conocimientos específicas del sector (ej. programación de sistemas TIC, antropología, o agricultura ecológica). En promedio, dos terceras partes de los conocimientos y habilidades demandadas por las vacantes en Colombia son conocimientos específicos del sector.

Tabla 2. Resumen de conocimientos y habilidades ofrecidos y demandados

Conocimiento/ Habilidad	Nivel de reutilización	Demanda laboral				Oferta educativa			
		Promedio	Promedio condicionado	Min.	Max.	Promedio	Promedio condicionad	Min.	Max.
Conocimientos	Tranversales	0,2	1,1	0	7	0,8	1,4	0	7
	Intersectoriales	0,8	1,9	0	20	14,8	14,8	0	79
	Específicos del sector	3,1	3,1	0	21	7,8	7,8	2	36
	Específicos de la ocupación	0,1	1	0	4	1	1,8	0	11
Habilidades	Tranversales	0,0	1	0	4	0	1	0	2
	Intersectoriales	0,2	1,6	0	16	0,1	2,9	0	21
	Específicos del sector	0,1	1,4	0	10	0,1	1,5	0	19
	Específicos de la ocupación	0,0	1	0	5	0	1,3	0	4
Total		4,6	4,6	0	39	24,7	24,7	2	129

Nota: Esta tabla presentan los promedios, el mínimo y el máximo de habilidades detectadas en las matrices de extracción de habilidades de vacantes y de planes de estudio según la clasificación de conocimientos y habilidades y el nivel de reutilización de los mismos. Las columnas promedio presenta el promedio considerando todas las vacantes y planes en las cuales no se detectaron habilidades, es decir, contabiliza los ceros de las matrices de extracción de habilidades. Las columnas promedio condicionado presentan el promedio solo de aquellas observaciones que detectaron al menos un (1) conocimiento o habilidad del nivel de reutilización analizado.

Sin embargo, al analizar la oferta de conocimientos y habilidades de la educación superior se observa que la principal oferta corresponde a conocimientos intersectoriales (ej. química, derecho civil, o diseño industrial). Aproximadamente 60 % de los conocimientos y habilidades enseñadas en la educación superior son conocimientos intersectoriales. Luego de este tipo de conocimientos, lo más enseñado por los programas son los conocimientos específicos del sector (31 %). Este análisis descriptivo permite evidenciar un desajuste entre los conocimientos que ofrecen los programas de educación superior y los que demandan los empleadores. Sin embargo, no permite concluir que la oferta educativa no sea adecuada para la demanda laboral, ya que,

mientras los programas ofrecen en promedio 25 conocimientos y habilidades, los empleadores solo demandan 5. Adicionalmente, se debe señalar que el promedio de la brecha de habilidad en los 3 años de estudio es de 83,45 %, siendo la brecha mínima de 43,73 % y la máxima de 100 %. Estas estadísticas dan una idea más aterrizada de la existencia de un fuerte desajuste entre la oferta educativa y la demanda laboral, pues, a pesar de que el sistema educativo ofrece 5 veces más habilidades de las que demanda el mercado laboral, las habilidades ofrecidas no corresponden a las demandadas.

5.2 Modelo

A través del modelo propuesto en la sección 4.2 es posible analizar la correlación, más no la causalidad, entre la brecha de habilidades y la probabilidad de vinculación laboral formal de los egresados de la educación superior en Colombia. Los resultados de las diferentes especificaciones presentadas en la tabla 3 muestran que un aumento en la brecha de habilidades, disminuye la probabilidad promedio de que los egresados de educación superior se vinculen al mercado laboral formal (en las 13 ciudades de la muestra), siendo esta correlación negativa consistente con el comportamiento teórico esperado. En el modelo M5, que es la especificación con todos los controles, se observa que por un aumento de un punto porcentual en la brecha de habilidades se reduce en 0,1484 puntos porcentuales la tasa de vinculación formal de los egresados.

Adicionalmente, los diferentes modelos estimados muestran que los egresados de instituciones con mejores resultados en los puntajes globales estandarizados de las pruebas Saber Pro y TyT tienen mayor probabilidad de vincularse al mercado formal. Sin embargo, es un efecto que desaparece al introducirse efectos fijos de municipio pues, al introducir este control la significancia del puntaje global desaparece. Este cambio puede deberse a un efecto del tamaño del mercado, ya que, al analizar las 13 ciudades¹⁹ incluidas en la muestra se observa que aquellos egresados que estudiaron en ciudades intermedias tienen menor probabilidad de vincularse formalmente al mercado laboral comparados con egresados en las 3 ciudades principales del país (Bogotá, Medellín y Cali).

También es relevante señalar que no existe un efecto de tiempo en la mayor o menor probabilidad de vincularse formalmente para los egresados de la educación superior, pero existe un efecto diferenciado según la área del conocimiento del programa de egreso. En el modelo M2 la dummy prueba específica que controla si los estudiantes de un programa presentan o no pruebas específicas en las pruebas Saber PRO y TyT muestra signo positivo, es decir, los egresados que presentan pruebas específicas tienen mayores probabilidades de vincularse formalmente. Esto puede deberse a un efecto de señalización, solo los programas que enseñan ciertos tipos de conocimientos presentan pruebas específicas, pues recordemos que las pruebas específicas son elección de los programas, pueden no presentarla, presentar una o hasta máximo 3 pruebas. En los modelos M3 a M5 la significancia de la dummy prueba específica desaparece por la inclusión del área del conocimiento en que se clasifica el programa, este zoom al interior de áreas de conocimiento permitió observar que, acorde con la literatura, el área de ingeniería, arquitectura, urbanismo y afines es la de mayor vinculación laboral formal para los egresados en el mercado

¹⁹Las ciudades incluidas en la muestra son: Medellín, Barranquilla, Bogotá, Cartagena, Tunja, Manizales, Popayán, Valledupar, Santa Marta, Cúcuta, Pereira, Bucaramanga y Cali

laboral colombiano, excepto por el área de las ciencias de la salud.

Tabla 3. Modelos lineales

Variable	TasaVinculación M1	TasaVinculación M2	TasaVinculación M3	TasaVinculación M4	TasaVinculación M5
Brecha de habilidad	-0.2755*** (0.0336)	-0.2406*** (0.0337)	-0.1714*** (0.0368)	-0.1711*** (0.0367)	-0.1484*** (0.0332)
año					
2016	(base)	(base)	(base)	(base)	(base)
2017	0.4016 (0.3351)	0.2947 (0.3303)	-0.0294 (0.3387)	-0.0075 (0.3413)	-0.0364 (0.3392)
2018	-0.0689 (0.3656)	-0.0176 (0.3621)	-0.3774 (0.3721)	-0.6315 (0.5762)	-0.9123 (0.5709)
Puntaje global Saber		1.8734*** (0.5086)	2.9581*** (0.4996)	2.9678*** (0.4994)	0.3068 (0.5118)
Prueba específica		3.6421*** (0.5815)	-0.4276 (0.6291)	-0.4384 (0.6282)	-0.0566 (0.6105)
Sector		2.3742*** (0.6927)	2.4480*** (0.6394)	2.4425*** (0.6388)	0.5471 (0.6192)
Grupo Ocupación			-2.2724 (2.7550)	-2.1712 (2.7628)	-1.1437 (2.5261)
Nivel Formación			-3.8207 (3.4463)	-3.9150 (3.4473)	-3.3843 (3.0524)
Modalidad			3.5649*** (1.1993)	3.5634*** (1.1990)	1.2517 (1.1718)
Egresados ocupación				-0.000009 (0.000017)	-0.000016 (0.000017)
constante	100.7432*** (2.6825)	92.8027*** (2.8396)	94.0820*** (3.9346)	94.2633*** (3.9345)	96.8763*** (3.6150)
Área del conocimiento	No	No	Si	Si	Si
Municipio	No	No	No	No	Si
N	4912	4770	4770	4770	4770
sigma_u	11.7755	11.3494	10.2947	10.2453	9.2639
sigma_e	9.1133	8.8829	8.8829	8.8830	8.8830
rho	0.6254	0.6201	0.5732	0.5708	0.5209

leyenda: * p<0.1; ** p<0.05; *** p<0.01

Nota: Esta tabla presentan los resultados de las estimaciones de mínimos cuadrados ordinarios con errores robustos a heterocedasticidad de las diferentes especificaciones analizadas. En todos los modelos la variable dependiente es la tasa de vinculación formal por programa del observatorio laboral para la educación. Adicionalmente, todos los modelos incluyen la brecha de habilidades y las dummies de años como controles, los demás controles varían a través de las especificaciones. El puntaje global saber corresponde al puntaje global estandarizado promedio de los estudiantes de un programa académico. Prueba específica es una dummy que indica si el programa presenta o no pruebas específicas en el Saber Pro y TyT. Sector indica si el programa se ofrece en una institución pública o privada. Grupo Ocupación es una dummy que indica si el programa corresponde a ocupaciones directivas o profesionales (CIUO 1 y 2) o a ocupaciones técnicas, profesionales de nivel medio, personal de apoyo o trabajadores de servicios (CIUO 3, 4 y 5). Nivel de Formación es una dummy que indica si el programa es universitario o técnico/tecnológico. Modalidad indica si los cursos del programa son a distancia tradicional/virtual o presenciales. Egresados ocupación corresponde al número de egresados de los programas agrupados según la CIUO a 2 dígitos. Área del conocimiento es medida acorde a la clasificación que hace el SNIES acorde con la clasificación del ICFES-Saber Pro y TyT. Municipio es un efecto fijo de lugar (con códigos divipola).

Deming and Noray (2020) encuentran que las carreras STEM (ciencias, matemáticas e ingenierías) tiene mayor facilidad para el tránsito de la educación superior hacia el mercado laboral. Sin embargo, en el contexto colombiano donde el ICFES y el registro SNIES cuentan como áreas separadas las ciencias y matemáticas de las ingenierías se observa que, los datos colombianos son consistentes con el hallazgo de Deming and Noray (2020) en cuanto a ingenierías pero no en ciencias y matemáticas, lo cual puede explicarse por el bajo desarrollo de ciencias básicas (matemáticas, física, química, etc.) en el país, tanto en el sector público como en el privado. El resultado de las ciencias de la salud también puede explicarse por el contexto regulatorio colombiano en el sector salud, pues es obligación para las EPS e IPS que sus trabajadores de la salud (por contrato laboral o prestación de servicios) contribuyan al régimen de seguridad social, es decir, que coticen a salud y pensión y, por tanto, existen altas tasas de vinculación formal medida como pago a salud y pensión.

6 Conclusiones

Esta investigación comenzó como un cuestionamiento sobre la importancia de las habilidades para encontrar trabajo formal en Colombia, preguntándose específicamente ¿Cómo afecta la brecha de habilidades entre la demanda laboral y la oferta de educativa a la tasa de vinculación laboral formal de los egresados de la educación superior? los análisis descriptivos y el modelo estimado permiten concluir que, para la muestra analizada (en 13 ciudades), encontrar trabajo formal es parcialmente cuestión de habilidades, es decir, que a mayor (menor) brecha de habilidad entre lo aprendido en el sistema de educación superior y lo demandado por el mercado laboral, menor (mayor) será la probabilidad de un egresado de cotizar al sistema de seguridad social, o, lo que es equivalente para un recién egresado, de vincularse formalmente al mercado laboral.

Con una brecha de habilidad mínima estimada de 43,73 %, una máxima de 100 % y un promedio de 83,45 %, los datos de la oferta educativa de programas de pregrado muestran un amplio desequilibrio respecto a la demanda del mercado laboral que ayuda a entender la dificultad de los jóvenes (sin experiencia) para encontrar trabajo, lo que se refleja en las altas tasas de desempleo que enfrentan los jóvenes en Colombia. En términos descriptivos, se encontró que mientras los programas de pregrado enseñan fundamentalmente conocimientos intersectoriales, es decir, aquellos útiles para ocupaciones en diferentes sectores de la economía, el mercado demanda principalmente conocimientos específicos de los sectores, es decir, aquellos útiles para diferentes ocupaciones dentro de un mismo sector de la economía. Dado lo anterior, las universidades deben atender de manera efectiva este desequilibrio para mejorar la empleabilidad de los jóvenes recién egresados. La forma de poder atender este desequilibrio entre la oferta y demanda de habilidades es a través de actualizaciones constantes de los planes curriculares, por ejemplo, manteniendo algunas asignaturas como pilares básicos de la formación disciplinar e implementando algunas cátedras complementarias que semestral o anualmente actualicen sus contenidos curriculares acorde a las nuevas demandas del mercado laboral, lo cual implica un monitoreo constante de las vacantes existentes en el mercado por parte de las universidades.

Cabe señalar que, la presente investigación presentó una serie de dificultades y limitaciones entre las que destacan: 1) la falta de datos previos sobre planes curriculares de educación superior

y, en general, de información sobre la oferta de habilidades en la educación superior y, 2) los errores introducidos por el reconocimiento óptico de caracteres y por la técnica estándar de diccionario para la indexación de habilidades (la clasificación ESCO para este caso). Ante la primera limitación se requirió priorizar los datos a analizar con el fin de brindar el mejor panorama posible del sistema de educación superior a nivel de pregrado, así, se priorizaron 13 ciudades que acumulan dos terceras partes de los 7422 programas activos del país, recopilándose información de un tercio de los programas. La segunda limitación requirió análisis manuales de depuración para la corrección y/o eliminación de las potenciales fuentes de error de medición con el fin de minimizarlo y brindar los resultados más certeros posibles.

Sin embargo, dichas limitaciones también se presentan como futuras líneas de investigación. Por un lado, es posible ampliar la muestra de programas para incorporar más programas de pregrado en ciudades no priorizadas en esta investigación y, para incluir también programas de posgrado con el fin de identificar los desequilibrios de habilidades en todo el sistema de educación superior. Por otro lado, es posible implementar nuevas y mejores técnicas de detección de habilidades, por ejemplo, usar técnicas vectoriales de similitud de textos usando las descripciones de las habilidades y conocimientos de la ESCO y no solo el nombre de dichas habilidades y conocimientos. También, es relevante profundizar en técnicas de reconocimiento óptico de caracteres que cometan menos errores y, por tanto, requieran menos supervisión humana manual.

Una tercera línea de investigación que se desprende de las limitaciones de este trabajo es buscar fuentes alternativas de aproximarse a la oferta de habilidades en el sistema educativo. Si bien los planes curriculares de los programas demostraron ser una proxy útil para los fines de esta investigación, no permiten el análisis granular a nivel de individuo pues, lo que se enseña en los programas no necesariamente es aprendido por todos los estudiantes ni en los mismos grados de comprensión. Así, buscar fuentes que permitan, por ejemplo, conocer las hojas de vida de los desocupados del mercado laboral permitiría realizar análisis de las habilidades efectivamente ofertadas en el mercado laboral y entender mejor aún el mismatch entre oferta y demanda en dicho mercado.

Referencias

- Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.
- Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of economic perspectives*, 25(3):153–72.
- Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.
- Atalay, E., Phongthientham, P., Sotelo, S., and Tannenbaum, D. (2017). The evolving us occupational structure. *Washington Center for Equitable Growth Working Paper*, 12052017.
- Carneiro, P., Crawford, C., and Goodman, A. (2007). The impact of early cognitive and non-cognitive skills on later outcomes.
- DANE (2021). Información histórica del mercado laboral: marco 2005.
- David, H., Katz, L. F., and Kearney, M. S. (2006). The polarization of the us labor market. *American economic review*, 96(2):189–194.
- Deming, D. J. and Noray, K. (2020). Earnings dynamics, changing job skills, and stem careers.
- Desjardins, R. and Rubenson, K. (2011). An analysis of skill mismatch using direct measures of skills.
- Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, 24(3):411–482.
- Hershbein, B. and Kahn, L. B. (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review*, 108(7):1737–72.
- ManpowerGroup (2021a). Employment outlook survey q3 2021 - colombia results.
- ManpowerGroup (2021b). Employment outlook survey q3 2021 - global results.
- Medina, C. and Posso, C. (2018). Cambio técnico y polarización en el mercado laboral. evidencia para colombia. *El trimestre económico*, 85(338):365–410.
- Modestino, A. S., Shoag, D., and Ballance, J. (2016). Downskilling: changes in employer skill requirements over the business cycle. *Labour Economics*, 41:333–347.
- Mortensen, D. T. (1970). Job search, the duration of unemployment, and the phillips curve. *The American Economic Review*, pages 847–862.
- Sepulveda, C., Gallego, J., Cardenas, J., Sarango, J., Roper, S., and Alianza, E. (2020). Empleabilidad e informalidad: un análisis del mercado laboral juvenil para 5 países latinoamericanos. Technical report.

Todd, P. E. and Zhang, W. (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative Economics*, 11(1):231–275.

Legislación colombiana sobre educación superior

Congreso de la República. (1992). Ley de educación superior: Ley 30 de Diciembre 28 de 1992. link de consulta: [Ley de educación superior](#)

Congreso de la República. (1994). Ley general de educación: Ley 115 de Febrero 8 de 1994. link de consulta: [Ley general de educación](#)

Congreso de la República. (2002). Ley de educación superior en las modalidades de formación técnica profesional y tecnológica: Ley 749 de Julio 19 de 2002. link de consulta: [Ley de educación superior en modalidades técnica y tecnológica](#)

MinEducación. (2019). Educación inicial. Link de consulta: <https://www.mineducacion.gov.co>

links de consulta de bases de datos

[Base SNIES de programas](#)

[Base SNIES de estudiantes graduados](#)

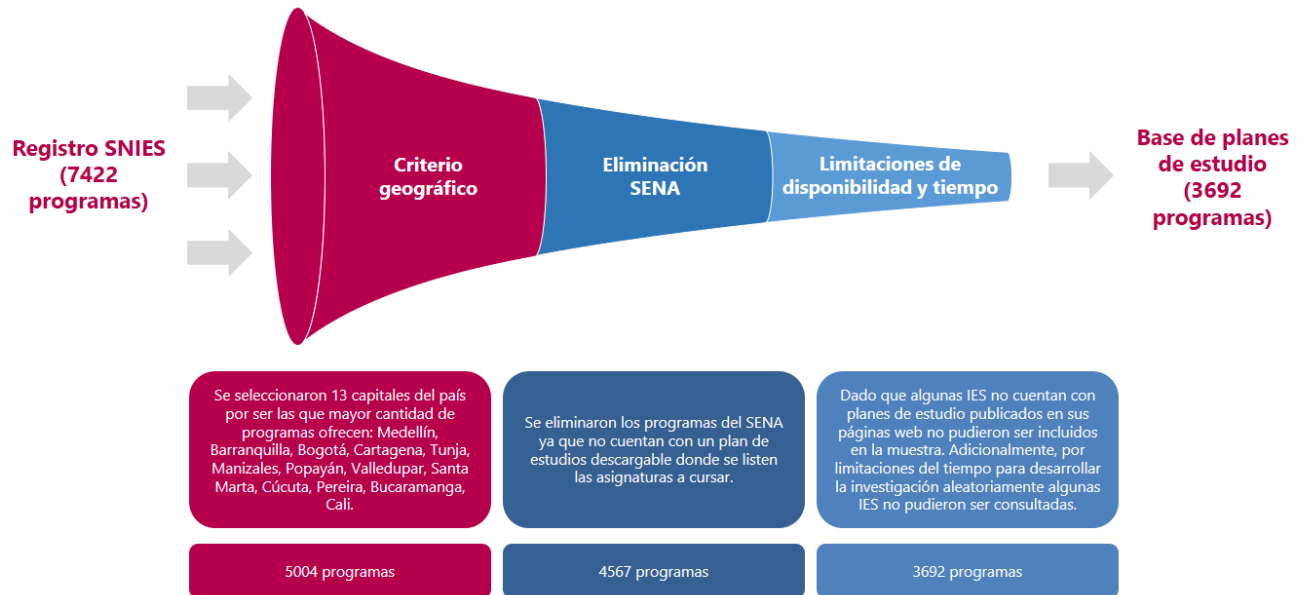
[Base OLE de tasa de cotización por programa](#)

[Clasificación ESCO versión 1.0.8 en español](#)

Las bases de vacantes fueron solicitadas al Servicio Público de Empleo bajo un acuerdo de uso exclusivo para esta investigación, por tanto, no existe un link público de consulta, para consultar dichas bases se debe elevar una solicitud al SPE. La base de planes de estudio fue una construcción propia para esta investigación, por tanto, tampoco existe un link público de consulta, para acceder a dicha base se requiere el consentimiento expreso del autor.

Anexos:

Anexo 1: Selección de la muestra de planes de estudio



Nota: Dado el criterio geográfico, la muestra seleccionada es urbana. Adicionalmente, la eliminación de los programas del SENA debido a la no disponibilidad de un plan de estudios implica la eliminación del 13,32 % del total de programas del país y el 8,66 % de los programas de las 13 capitales. Cabe resaltar que los programas del SENA en la base del SNIES son de carácter oficial y tecnológico.

Anexo 2: Caracterización del SNIES, muestra planes de estudio y planes no incluidos en la muestra.

Características	SNIES		Muestra planes de estudio		No incluidos en la muestra		Diferencia de proporciones (muestra - no muestra)		
	Nivel	%	Nivel	%	Nivel	%	p.p.	Dif. significativa	
Número de programas	7422	100	3692	100	3730	100	-	-	
Sector	Público	3178	42,82	907	24,57	2271	60,88	-36,31	Si
	Privado	4244	57,18	2785	75,43	1459	39,12	36,31	Si
Nivel de formación	Técnica	687	9,26	295	7,99	392	10,51	-2,52	Si
	Tecnológica	2301	31,00	685	18,55	1616	43,32	-24,77	Si
	Universitaria	4434	59,74	2712	73,46	1722	46,17	27,29	Si
Modalidad	Presencial	6626	89,32	3211	87,00	3415	91,63	-4,63	Si
	Distancia tradicional	313	4,22	131	3,55	182	4,88	-1,33	Si
	Distancia virtual	479	6,45	349	9,46	130	3,49	5,97	Si
Alta calidad	Si	1461	19,68	1062	28,76	309	10,70	18,06	Si
	No	5930	79,90	2625	71,10	3305	88,61	-17,51	Si
	No aplica	31	0,42	5	0,14	26	0,70	-0,56	Si

Nota: La modalidad dual (4 programas) no fue tomada en cuenta para el cálculo de las estadísticas de modalidad, por tanto, la suma de los niveles difiere del total. La significancia de las diferencias fue calculada a través de una prueba de hipótesis sobre la diferencia de proporciones poblacionales con un nivel de confianza del 95 %, donde la hipótesis nula es que la diferencia de proporciones poblacionales es cero (0). Bajo las diferentes características analizadas, los programas de la muestra son diferentes a los programas no incluidos en la muestra, estas diferencias se explican por: 1) la concentración urbana de la muestra (13 capitales) que aumenta la probabilidad de que un programa consultado aleatoriamente cuente con acreditación de alta calidad, y 2) la eliminación de los programas del SENA de la muestra, pues esto representa el 40 % de los programas públicos no incluidos en la muestra y el 60 % de los programas tecnológicos no incluidos en la muestra.